# The Sleep Onset Transition: a Connectivity Investigation Built on EEG Source Localization

---

**Dissertation**

**zur**

**Erlangung der naturwissenschaftlichen Doktorwürde**

**(Dr. sc. nat.)**

**vorgelegt der**

**Mathematisch-naturwissenschaftlichen Fakultät**

**der**

**Universität Zürich**

**von**

Antonio Fernández Guerrero

**aus**

Spanien

Promotionskommission

Prof. Dr. Peter Achermann (Vorsitz und Leitung der Dissertation)

Prof. Dr. Jean-Marc Fritschy

Prof. Dr. Dr. med. Klaas Enno Stephan

**Zürich,** 2018

**Acknowledgements**

I want to credit special thanks to the following people for helping me, in one way or another, to make possible the final redaction of this thesis: Prof. Dr. Alexander Borbély (for helping with the corrections on most chapters of this thesis and offering additional insightful remarks), Dr. Benjamin Stucky and Dr. Dario Dornbierer (for helping with the German translation of the abstract), Mr. Ian Clark (for helping with the writing style and use of correct grammar) and Dr. Leila Tarokh (for her fruitful feedback on chapter 3). Special thanks to Prof. Dr. Peter Achermann, whom I consider a true mentor, for steadily supporting me in numerous ways during all these years and, most importantly, having being the most crucial figure in my formation as a neuroscientist.

## Table of Contents

**Summary** ..... vi

**Zusammenfassung** ..... xi

1. **Sleep, its potential functions and sleep onset** ..... **1**

1.1. Sleep phases and principal neurophysiological mechanisms ..... 1

1.1.1. NREM sleep ..... 4

1.1.2. REM sleep ..... 7

1.2. Biological functions and evolutionary theory of sleep ..... 10

1.2.1. Energy conservation and allocation theory ..... 12

1.2.2. Glymphatic system and immune boosting ..... 15

1.2.3. Brain plasticity, memory and learning ..... 18

1.3. Sleep onset ..... 22

1.3.1. Sleep onset transition: neurophysiological aspects ..... 22

1.3.2. Sleep onset transition: insights from latest quantitative research ..... 25

2. **Brain connectivity and EEG source localization** ..... **39**

2.1. Brain connectivity: introductory overview ..... 39

2.1.1. Structural connectivity ..... 42

2.1.2. Functional connectivity ..... 44

2.1.3. Effective connectivity ..... 46

2.2. Source localization with focus on low resolution electromagnetic tomography (LORETA) ..... 50

2.3. Applications of LORETA to sleep ..... 56

3. **Brain dynamics during the sleep onset transition: an EEG source localization study** ..... 72

3.1. Introduction ..... 73

3.2. Methods ..... 75

3.3. Results ..... 77

3.3.1. Temporal evolution of delta activity at sleep onset ..... 77
3.3.2. Temporal evolution of sigma (spindle) activity at sleep onset ..... 84
3.3.3. Temporal evolution of alpha activity at sleep onset..... 91
3.3.4. Temporal evolution of beta activity at sleep onset..... 92
3.3.5. Temporal evolution of theta activity at sleep onset ..... 92
3.4. Discussion..... 93
3.4.1. Delta activity ..... 93
3.4.2. Sigma activity..... 95
3.4.3. Other frequency bands..... 97
3.5. Limitations..... 98
3.6. Conclusions..... 98
3.7. Supplementary Material ..... 99
3.7.1. Supplementary Methods..... 99
3.7.2. Supplementary Results ..... 100
3.7.2.1. Temporal evolution of apha activity at sleep onset ..... 100
3.7.2.2. Temporal evolution of beta activity at sleep onset ..... 104
3.7.2.3. Temporal evolution of theta activity at sleep onset ..... 107
4. **Assessing intracortical causal information flow of oscillatory activity at the sleep onset transition with the isolated effective coherence (iCOH)** ..... 111
4.1. Introduction..... 112
4.2. Methods ..... 114
4.2.1. Data description and preprocessing..... 114
4.2.2. Isolated effective coherence (iCOH)..... 116
4.2.3. Selection of ROIs ..... 117
4.2.4. Presentation of results and statistical analyses..... 119

4.3. Results ..... 120
4.3.1. iCOH spectra ..... 120
4.3.1.1. Baseline condition ..... 120
4.3.1.2. Recovery condition ..... 122
4.3.1.3. Statistical contrast between baseline and recovery conditions ..... 123
4.3.2. iCOH of specific frequency bands ..... 124
4.3.2.1. Delta band ..... 126
4.3.2.2. Sigma band ..... 127
4.3.2.3. Additional frequency bands ..... 128
4.4. Discussion ..... 130
4.4.1. Relevant spectral features of iCOH ..... 130
4.4.1.1. Baseline ..... 130
4.4.1.2. Recovery ..... 131
4.4.1.3. Recovery versus baseline ..... 132
4.4.2. Topographical properties and neurobiological interpretation ..... 132
4.5. Supplementary Material ..... 140
4.5.1. Brief explanation of iCOH formula and characteristics ..... 140
4.5.2. Theoretical aspects ..... 142
4.5.3. Statistics ..... 144
4.5.4. Extended discussion on additional frequency bands ..... 145
5. **Summary of connectivity methods and their applications in sleep research** ..... 150
5.1. Introduction ..... 151
5.2. Functional connectivity (FC) ..... 158
5.2.1. Information theory ..... 158
5.2.2. Mutual information (MI) ..... 159
5.2.2.1. Cross-mutual information (CMI) ..... 161
5.2.2.2. Auto-mutual information (AMI) ..... 162
5.2.3. Partial information (PI) ..... 164

5.2.3.1. Partial mutual information (PMI) ..... 165
5.2.4. Applications to sleep EEG recordings..... 172
5.3. Effective connectivity..... 174
5.3.1. Transfer entropy ..... 174
5.3.2. Granger causality..... 175
5.3.3. Applications to loss of consciousness (LOC) ..... 177
5.3.4. Directed transfer function (DTF), partial directed coherence (PDC) and isolated effective coherence (iCOH)..... 180
5.3.5. Applications to sleep studies and related areas ..... 190
6. **Discussion**..... 208
6.1. Limitations: practical and theoretical ..... 208
6.1.1. Number of EEG recording channels and relationship with LORETA ..... 208
6.1.2. Number of participants and statistics ..... 209
6.1.3. Use of Magnetoencephalography (MEG) and comparison with EEG ..... 211
6.1.4. iCOH: criticism and effects when combined with LORETA ..... 212
6.2. Future directions to explore..... 216
6.2.1. Sleep and creativity ..... 217
6.2.2. Sleep and consciousness..... 221
6.2.3. Sleep and personality research ..... 230
6.3. Concluding remarks..... 239
**List of abbreviations** ..... 241
**References**..... 245
**Quotations**..... 264
**Curriculum vitae**..... 267

## Summary

*"We are such stuff as dreams are made on, and our little life is rounded with a sleep."* William Shakespeare[0]

The study of sleep is of central interest, both for theoretical reasons as a basic scientific question, as the purpose of sleep is still partly unknown, and for clinical applications in the improvement of human health. From a health perspective, sleep is of importance, not only for its role in basic cognitive functioning, such as memory, but also for the quality of life in general. In this thesis, we have focused on the period of the sleep onset (SO) transition, having a special interest as it covers a kind of "phase-transition" of the brain, gradually shifting from a state of conscious wakefulness to unconscious disengagement from the external environment (sleep). Therefore, given the proximity between the topics of SO and consciousness, our research may also foster studies of the neurobiological basis of consciousness.

Chapter 1 contains a general theoretical introduction into the field of sleep, addressing questions such as the most relevant neurophysiological correlates of sleep, the principal characteristics of sleep stages, a possible evolutionary basis, and a concise discussion on different theories (e.g., energy savings, memory and learning) about its biological function(s). Previous studies on the neurobiological understanding of the SO transition are also summarized.

In chapter 2, the concept of connectivity in neuroscience is thoroughly explained, clarifying common misconceptions, and distinguishing between the three different connectivity subtypes (structural, functional, and effective). The chapter also provides the general basis for each subtype (including potential limitations in applications or other caveats). Additionally, an introduction into the inverse problem, source localization, i.e. estimating the cortical sources that generated the EEG recorded at the scalp, with special prominence given to LORETA (low resolution electromagnetic tomography) is presented. Finally, an overview of LORETA applied in sleep studies is provided, with an emphasis on the SO transition when possible.

To reach a more complete and integrated understanding of the SO process, we have combined two methodological approaches of computational neuroscience: source localization and effective connectivity (chapters 3 and 4, respectively). Firstly (chapter 3), we applied LORETA, a source localization tool which has the property of providing solutions for the estimation of the underlying neuronal generators of scalp recorded EEG activity with zero-error. LORETA is an exact method for reconstructing sources of brain activity from EEG recordings, although manifesting some blurring. Thus, LORETA allowed us to pinpoint the cortical brain areas exhibiting a guiding role during the SO transition: how they evolve with time and how they differ in intensity levels along different frequency bands. Significant changes in activity levels of the diverse cortical areas paralleling SO were tracked with statistical non-parametric mapping (a bootstrapping method based on surrogates). Statistical contrasts searching for changes with SO (defined as the first occurrence of stage 2 non-REM sleep) were performed for two conditions: a baseline night (normal sleep pressure, 16 h of wakefulness) and a recovery night (increased sleep pressure, 40 h of sustained wakefulness).The temporal evolution was tracked in 2-min intervals: one prior to SO, and five 2-min intervals following SO.

In particular, we followed the SO process in two major frequency bands: delta (0.5-5 Hz) and sigma (12-16 Hz). LORETA images showed that delta activity steadily increased with time in frontal regions, reaching a maximum at the end of the 10 min window. The steady raising of delta activity occurred in both conditions but was significantly steeper in recovery (as a result of increased homeostatic sleep pressure, which accelerated the temporal dynamics compared to baseline) and fitted well with an exponential increase. The brain areas showing an initial quicker involvement during the SO transition comprised mainly the frontal lobe, with central and posterior areas getting incorporated around 4-6 min in baseline and somewhat earlier in recovery due to higher sleep pressure. On the other hand, activity in the sigma band (12-16 Hz, related to spindle activity, particularly fast spindles), followed a time course that is analogous to an inverted U-shape, both in baseline and recovery conditions (although with some differences, attributed the increased homeostatic pressure). The

maximum of activity was located in the parietal cortex (in agreement with previous literature on fast spindle activity). The recovery condition displayed typical lower levels of spindle activity than baseline and the maximum occurred earlier, consistent with temporally accelerated dynamics. In general, the recovery condition manifests as a temporally accelerated version of baseline sleep in both frequency bands as a prominent hallmark of increased sleep pressure.

Further (chapter 4), we focused on effective connectivity at the level of the source space, as rendered by LORETA. We used the lagged isolated effective coherence method, which is a linear method to study effective connectivity. This technique reconstructs spectral characteristics of the transmission of information between nodes, distinguishes between the two possible directions in a given pair and assesses without distortion the connection strength between a pair of nodes (discounting indirect connections, a caveat of some other methods). We defined as regions of interest nine nodes comprising the default mode network (DMN; including, among others, the posterior cingulate cortex, and the middle prefrontal cortex), dorsolateral prefrontal cortex and hippocampus. We obtained a posterior to anterior decoupling of the DMN in the low frequency range reflecting the progressive disengagement from the external environment with SO. The posterior cingulate cortex (PCC) emerged as playing a major role in the unfolding of the SO transition, guiding the other nodes, particularly, in the delta-theta range. Furthermore, the midcingulate cortex appeared as a principal cortical relay station of the thalamus in spindle synchronization throughout the cortex. The hippocampus also appeared to be involved in the coordination of the SO process, showing significant effective connectivity values in the theta and sigma bands. Under conditions of sleep deprivation, posterior-anterior decoupling of the DMN (specifically, between the posterior and anterior cingulate and adjacent frontal regions) in the low frequency range and sigma band occurred exhibiting overall smaller effective connectivity changes. Lower overall cortical connectivity in the recovery condition was present due to sustained wakefulness, so the transition exhibited a lower reduction than in baseline still leading to a disconnection between the major nodes of the DMN. Finally, we also observed

hindered transmission in the spindle range of cortico-cortical interactions after sleep deprivation (in addition to their diminished activity).

In addition, in chapter 5 we reviewed thoroughly suitable methods for the study of brain connectivity (except structural connectivity, related to white matter bundles). A selection of prominent methods, covering both functional (e.g., coherence, cross mutual information) and effectivity connectivity (e.g., Granger causality, partial directed coherence), are first described in detail and then shown how they have been applied in sleep research. We have tried to clearly distinguish (as sometimes it is mixed or confused in the literature) functional from effective connectivity, their main differences and best occasion for application of each of them (according to the nature of the problem and specific questions addressed). Furthermore, we have put an emphasis on the highly complex notion of "causality", trying to clarify this very concept (at least, to some extent) and convey its most important features, and how they relate to the family of effective connectivity methods, which are supposed to characterize causal relationships. We have incorporated in our review both linear and non-linear connectivity methods. In order to facilitate a fluent reading and reach a greater audience of researchers, we provide explanations in intuitive terms, nonetheless, without neglecting the required rigor in this abstract, yet fundamental, topic.

Finally, chapter 6 exposes the concluding remarks and both practical (e.g., moderately low number of subjects and EEG channels) and theoretical (e.g., refinements to the iCOH technique) limitations that were faced and how they could be improved in future studies. It is argued, first, that the fading of consciousness during SO can be understood by reduced small-world topology and higher modularity, secondly, that the PCC plays a pivotal role in the disconnection of the DMN, and finally, that spindle activity is anti-correlated with delta activity, showing a curvilinear evolution with the midcingulate cortex acting as a prominent cortical relay station for spindle diffusion. Relationships between sleep and other major neuroscientific areas, including inter-individual differences in personality and creativity levels are also discussed.

## Zusammenfassung

Die Erforschung des Schlafes ist von zentralem Interesse, sowohl aus theoretischen Gründen, da die Funktion des Schlafes teilweise immer noch unbekannt ist, als auch aus klinischer Sicht um die menschliche Gesundheit zu verbessern. Der Schlaf hat einen grossen Einfluss auf die Gesundheit, nicht nur aufgrund seiner Funktion was grundlegende kognitive Funktionen wie das Erinnerungsvermögen anbelangt, sondern auch für das generelle Wohlbefinden. In dieser Arbeit haben wir uns auf die Einschlafphase fokussiert. Diese kann als sukzessiver Phasenübergang des Gehirnes vom bewussten Wachzustand hin zu einem von der Umgebung losgelösten unbewussten Zustand (Schlaf) verstanden werden. Hinsichtlich der zentralen Rolle des Bewusstseins während des Einschlafens, kann unsere Forschung auch einen Beitrag für das Verständnis der neuronalen Grundlagen des Bewusstseins leisten.

Kapitel 1 enthält eine generelle theoretische Einführung in die Schlafforschung. Dabei werden grundlegende Konzepte wie z.B. die wichtigsten neurophysiologischen Korrelate von Schlaf, die wesentlichen Merkmale der Schlafstadien, mögliche evolutionäre Erklärungen des Phänomen Schlaf und eine Zusammenfassung verschiedener Theorien seiner biologischen Funktionen (z.B. Energie-Einsparung, Erinnerungsvermögen und Lernen) erörtert. Zudem werden die Resultate früherer Studien, welche sich mit den neurobiologischen Grundlagen des Einschlafens befasst haben, zusammengefasst.

Im Kapitel 2 wird das Konzept der *neurowissenschaftlichen Konnektivität* sorgfältig erörtert. Unter anderem werden weitverbreitete Missverständnisse aufgeklärt und zwischen den drei Subtypen der Konnektivität (strukturell, funktionell, effektiv) unterschieden. Auch die Grundlagen für jeden dieser Subtypen (einschliesslich der möglichen Anwendungslimitationen oder anderer Vorbehalte) werden aufgezeigt. Des Weiteren wird eine Einführung in das *inverse Problem*, die Quellenlokalisation gegeben. Die Quellenlokalisation gibt eine Schätzung der möglichen kortikalen Quellen, welche das auf der Kopfhaut aufgezeichnete EEG generieren. Dazu wird eine häufig

verwendete Software, LORETA (low-resolution electromagnetic brain tomography), vorgestellt. Es wird zudem eine Übersicht der Studienlage gegeben, welche mithilfe von LORETA den Einschlafprozess untersucht haben.

Um ein ganzheitlicheres Verständnis über den Einschlafprozess zu gewinnen, wurden zwei methodologische Herangehensweisen aus der Neuroinfomatik kombiniert, nämlich die Quellenlokalisation und die effektive Konnektivität (Kapitel 3 und 4). Als Erstes (in Kapitel 3) wurde LORETA eingesetzt, eine Quellenlokalisierungsmethode welche es erlaubt, aufgrund des auf der Kopfhaut gemessenen EEGs die zugrundeliegenden neuronalen Generatoren zu berechnen. LORETA ist eine exakte Methode, mit welcher basierend auf dem EEG Signal, die neuronalen Quellen der Gehirnaktivität rekonstruiert werden können, auch wenn diese eine gewisse Unschärfe aufweisen. Somit wurden mithilfe von LORETA diejenigen kortikalen Gehirnareale ermittelt, welche eine wichtige Rolle während des Einschlafprozesses spielen, wie sie sich diese über die Zeit hinweg verändern und wie sich ihre Intensität in den verschiedenen Frequenzbändern unterscheiden. Mithilfe von nicht-parametrischen statistischen Tests (eine auf Surrogaten basierte Bootstrap-Methode) wurden signifikante Veränderungen in den Aktivitäts-Niveaus verschiedener kortikalen Areale ermittelt, welche im Zusammenhang mit dem Einschlafprozess stehen. Statistische Veränderungen während des Einschlafens (definiert als das erste Auftreten von Stadium 2 des non-REM-Schlafs) wurden für zwei Bedingungen untersucht: eine Kontrollnacht (normaler Schlafdruck, 16 h Wachheit) und eine Erholungsnacht (erhöhter Schlafdruck, nach 40 h ununterbrochener Wachheit). Die zeitliche Entwicklung wurde in 2 Minuten Blöcken gemessen, ein Block vor dem Einschlafen, und danach fünf Blöcke nach dem Einschlafen.

In erster Linie wurde die Analyse des Einschlafprozesses auf zwei Frequenzbänder limitiert: dem Delta- (0.5 - 5 Hz) und Sigma-Band (12 – 16 Hz). Die LORETA-Analysen ergaben, eine über die Zeit stetige Zunahme der Delta-Aktivität in frontalen Regionen, welche nach 10 Minuten ihr Maximum erreichte. Diese stetige Zunahme der Delta-

Aktivität konnte in beiden Bedingungen beobachtet werden, sie war jedoch in der Erholungsnacht signifikant steiler. Der Grund dafür ist, dass der homöostatische Schlafdruck in der Erholungsnacht erhöht war, was zu einer Beschleunigung der zeitlichen Dynamik im Vergleich zur Kontrollnacht führte. Der Aktivitätsanstieg folgte einer exponentiellen Zunahme. Dabei zeigte der Frontallappen den schnellsten Aktivitätsanstieg, gefolgt vom zentralen und parietalen Kortex, welche in der Kontrollnacht nach etwa 4-6 Minuten und in der Erholungsnacht – aufgrund des höheren Schlafdrucks – eine frühere Mitbeteiligung zeigte. Dem gegenüber steht das Sigma-Band (12- 16 Hz, wird mit Spindelaktivität in Verbindung gebracht, insbesondere den schnellen Spindeln), dessen Zeitverlauf eher einer invertierten U-Form folgt, aber für beide Bedingungen ähnlich war (es gibt jedoch Unterschiede welche dem Schlafdruck zugeordnet werden können). Die maximale Aktivität wurde im parietalen Kortex lokalisiert (welches auch mit vorangehender Literatur über schnelle Spindelaktivität übereinstimmt). Die Erholungsnacht zeigte typischerweise tiefere Werte für die Spindelaktivität verglichen mit der Kontrollnacht, wobei das Maximum früher auftrat, was wiederum konsistent ist mit einer zeitlich beschleunigten Dynamik. Die Erholungsnacht zeigt, was diese beiden Frequenzbänder anbelangt, eine beschleunigte Dynamik der Kontrollnacht, welche sich auf den erhöhten Schlafdruck zurückführen lässt.

Weiterführend (in Kapitel 4) wurde der Fokus auf die Analyse der effektiven Konnektivität innerhalb des Quellenraums gelegt, welcher mithilfe von LORETA ermittelt wurde. Dabei wurde die Methode *der verzögert-isolierten effektiven Kohärenz (lagged isolated effective coherence)* angewendet. Es handelt sich um eine lineare Methode, mit deren Hilfe sich die effektive Konnektivität beschreiben lässt. Diese Technik rekonstruiert spektrale Charakteristiken der Informationsübertragung zwischen Knotenpunkten, unterscheidet zwischen den beiden möglichen Richtungen der Informationsübertagung zwischen Kontenpaaren und bewertet die Stärke der Konnektivität zwischen ihnen (wobei indirekte Verbindungen ignoriert werden, ein Problem das bei anderen Methoden auftreten kann). Für die Analysen wurden 9 Interessensregionen (ROIs) als Knoten gewählt, darunter ein

Ruhezustandsnetzwerk (englisch: default mode network (DMN), bestehend aus dem posterioren cingulären Kortex (PCC) und dem mittleren Präfrontalkortex (mPFC)), den dorsolateralen Präfrontalkortex (DLPFC) und dem Hippocampus. Dabei konnte eine posterior-anteriore Entkoppelung innerhalb des DMN in den tiefen Frequenzbereichen beobachtet werden, welche die zunehmende Entkopplung von der externen Umgebung während des Einschlafprozesses widerspiegelt. Dabei stellte sich der PCC als Hauptknoten während des Einschlafens heraus, welcher andere Knoten insbesondere im Bereich der Delta-Theta-Aktivität steuert. Ferner stellte sich der mittlere cinguläre Kortex (MCC) als kortikale Relaisstation des Thalamus bei der kortikalen Spindelsynchronisierung heraus. Zudem schien auch der Hippocampus – aufgrund signifikanter effektiver Konnektivitätswerte im Theta- und Sigmaband – eine Rolle bei der Koordination des Einschlafprozesses zu spielen. In der Schlafentzugsbedingung zeigte die Entkopplung innerhalb des DMN längs der posterior-anterioren- Achse (genauer, zwischen dem PCC und dem ACC und benachbarten frontalen Regionen) in den tiefen Frequenzbändern und dem Sigma-Band insgesamt kleinere effektive Konnektivitätsänderungen. Insgesamt war die kortikale Konnektivität in der Erholungsnacht kleiner aufgrund der vorangehenden ununterbrochenen Wachheit. Der Übergang wies eine tiefere Reduktion verglichen mit der Kontrollnacht auf, welche jedoch immer noch zu einer Entkoppelung zwischen den Hauptknoten innerhalb des DMN führte. Zuletzt wurde auch eine verminderte kortiko-kortikale Interaktion im Spindelfrequenzbereich nach dem Schlafentzug gefunden (zusätzlich zur verminderten Aktivität).

Des Weiteren wurden in Kapitel 5 geeignete Methoden für die Untersuchung der Gehirnkonnektivität begutachtet (ausgenommen der strukturellen Konnektivität basierend auf der weissen Materie). Eine Auswahl gängiger Methoden, welche funktionelle (z.B. Kohärenz, Cross Mutual Information) und effektive Konnektivität (z.B. Granger-Kausalität, partiell-ausgerichtete Kohärenz) beinhalten, wurden im Detail beschrieben und danach aufgezeigt, wie diese in der Schlafforschung Anwendung finden. Es wurde versucht die Unterschiede zwischen funktionaler und effektiver Konnektivität

aufzuzeigen, da diese zwei Formen der Konnektivität in der Literatur häufig vermischt oder verwechselt werden. Dabei wurden die Hauptunterschiede und die jeweils geeignetsten Anwendungsgebiete erörtert (je nach Fragestellung und der Natur des Problems). Weiterhin wurde viel Wert auf die komplexe Idee der «Kausalität» gelegt. Es wurde versucht, das Konzept möglichst klar darzustellen (jedenfalls bis zu einem gewissen Grade) und die wichtigsten Eigenschaften und ihren Bezug zur effektiven Konnektivität zu vermitteln (welche dafür vorgesehen sind Kausalzusammenhänge zu beschreiben). Wir haben in unserer Übersicht lineare und nicht-lineare Methoden der Konnektivitätsberechnung miteinbezogen. Um den Lesefluss zu erhöhen und mehr Wissenschaftler zu erreichen, wurden die Erklärungen intuitiv gehalten, ohne jedoch die notwendige Genauigkeit für dieses abstrakte, aber auch fundamentale Thema zu vernachlässigen.

Abschliessend werden im Kapitel 6 die Schlussbemerkungen und praktische (z.B. eher kleine Anzahl an Probanden und EEG-Kanälen) sowie theoretische (z.B. Verfeinerungen der iCOH Technik) Limitationen der Arbeit und mögliche Verbesserungsvorschläge für zukünftige Studien diskutiert. Es wird erstens argumentiert, dass die Abnahme des Bewusstseinsgrades während des Einschlafprozesses als eine Reduktion der «small-world topology» mit erhöhter Modularität verstanden werden kann. Zweitens wird erörtert, dass der PCC eine ausschlaggebende Rolle in der Entkopplung des DMN spielt, und schlussendlich, dass Spindelaktivität in einem antikorrelierten Verhältnis zur Delta-Aktivität steht, welche durch eine kurvilineare Entwicklung gekennzeichnet ist und dass der MCC als Relaisstation für die Spindeldiffusion fungiert. Zusammenhänge zwischen dem Schlaf und anderen bedeutenden neurowissenschaftlichen Forschungsgebieten, wie zum Beispiel interindividuelle Unterschiede in Persönlichkeit und Kreativität, werden auch angesprochen.

# 1. Sleep, its potential functions and sleep onset

## 1.1. Sleep phases and principal neurophysiological mechanisms

*"Sleep is the interest we have to pay on the capital which is called in at death; and the higher the rate of interest and the more regularly it is paid, the further the date of redemption is postponed."* Arthur Schopenhauer[1.1]

Sleep has been defined as "a reversible behavioral state of perceptual disengagement from and unresponsiveness to the environment" (Cartwright, 2001). As it is known, the central nervous system (CNS) exhibits, in very broad terms, three distinct neurobiological states, which do not mix simultaneously under normal conditions: wakefulness, non-rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep (Porkka-Heiskanen, 2013). Prominent exceptions to this general statement include lucid dreaming, which shows the most distinctive waking feature (i.e., a demonstrable state of consciousness) while set on a REM sleep background, and some rare parasomnias (e.g., sleep terrors and somnambulism, introducing waking behaviors during NREM sleep) (Hobson and McCarley, 2003, Hobson, 2009). A healthy subject would begin to sleep through NREM sleep (important exceptions hold for babies and narcoleptics), which can be further subdivided into three different states, based on specific characteristics assessed by the electroencephalogram: stages 1, 2 and 3 (Iber et al., 2007a).

Stage 1 is the lighter form of NREM sleep, directly succeeding wakefulness in the absence of brain pathologies (although there is no consensus whether stage 1 represents genuine sleep), stage 3 corresponds to deep sleep (where arousal thresholds to awaken a subject are maximum) and stage 2 (also called light sleep) is the first stage generally agreed to be true sleep (Picchioni et al., 2008). Throughout the night, NREM and REM sleep alternate with a period of approximately 90 min, with NREM sleep predominating during the first third of the night and REM sleep during the last third of sleep (Ogilvie, 2001). A rough statistic quantifying the percentage of time that young healthy humans spend on each stage renders these figures: 5% stage 1, 55% stage 2 and 15% on stage 3 (together constituting 75-

80% of time in NREM sleep) and 20-25% on REM sleep (Cartwright, 2001).

Based on neurophysiological recordings in both humans and animals, the regulation of sleep homeostasis has been successfully modeled in the two-process model of sleep regulation (Borbély et al., 2016). The model postulates that a homeostatic process (called Process S) interacts with a process controlled by the circadian pacemaker (Process C) (Borbély, 1982, Daan et al., 1984). Process C relates to circadian rhythm and the regulation of internal hormonal and neurological processes throughout the entire organism (e.g., cortisol levels in the blood or vigilance and alertness in the brain), fundamentally controlled by a master clock located in the suprachiasmatic nucleus of the hypothalamus (Daan et al., 1984, Borbély et al., 2016). Process S posits a continuous increase in sleep pressure or propensity with daily activities (especially high energy demanding tasks, either physically or mentally, e.g., sports or chess (Dolezal et al., 2017)) and time spent awake, which has to be later compensated by NREM sleep due to its role in sleep homeostasis (Cartwright, 2001). Sleep homeostasis implies that a prolonged period of wakefulness is subsequently followed by increased intensity of slow waves during deep sleep and possibly also a prolonged period of sleep, in a causal (not merely correlative) process, thus, paying the previously accumulated homeostatic “debt” (Porkka-Heiskanen, 2013). In this way, after sleep deprivation, both NREM and REM sleep increase in total duration (yet, the main compensatory mechanism in NREM sleep is still through increased intensity), although REM sleep only after selective REM sleep deprivation (although REM rebound can be delayed to subsequent nights or even not be present at all, and the rebound is smaller than expected) (Borbély and Achermann, 1999, Rodriguez et al., 2016).

An essential variable appearing in the mathematical modelling of Process S is slow wave activity (SWA), which dissipates exponentially throughout the course of the night during NREM sleep (Borbély, 1982, Daan et al., 1984, Borbély and Achermann, 1999). SWA is a variable that can be extracted from the electroencephalogram (EEG) through

signal processing methods (most commonly relying on either the Fourier or Hilbert Transform), quantifying power spectral density (a variable proportional to the square amplitude of a wave at a given frequency) in the delta band (range of 0.5-4.5 or 5 Hz) (Finelli et al., 2001a). Exceptions to a smooth exponential decay process include mental pathologies such as depression and narcolepsy, characterized by a very fragmented sleep and other major neurophysiological disruptions under baseline conditions (i.e., not previously reliant on sleep deprivation protocols, thus, masking the observed deficits) (Ogilvie, 2001). Mathematical modelling of Process S, even when simplified, allows to predict with notable accuracy the relative value of SWA at the time of sleep onset (SO) through an equation based on a saturating exponential function (i.e., asymptotically convergent) that uses prior total waking time as the input variable (Vassalli and Dijk, 2009, Rusterholz et al., 2010). In this way, SWA can be regarded as a mathematical biomarker proportional to sleep need and the homeostatic component of sleep described by Process S (De Gennaro et al., 2000a, Marzano et al., 2013). Thus, Process S, correlating with SWA (a marker which can be used to assess it), increases steadily during wakefulness (representing sleep debt) and declines exponentially during sleep (Borbély and Achermann, 1999). Furthermore, as Process S (in its interplay with Process C) declining during sleep, approaches a lower threshold, it triggers awakening, while increasing during waking, approaching the upper threshold triggers sleep (Borbély et al., 2016).

A very relevant biochemical substance closely related to Process S and sleep homeostasis is adenosine, a molecule resulting from the hydrolysis of adenosine triphosphate (ATP) (Cartwright, 2001, Garcia-Rill et al., 2006, Landolt, 2008). Continuous activity during wakefulness, being either mental or physical, demands large energy requirements to the cells, which are supplied by ATP; in turn, ATP produces adenosine after energetic depletion and phosphate breakdown (Vassalli and Dijk, 2009). As a result, adenosine levels built up steadily during wakefulness, particularly in the basal forebrain, where adenosine binds to A1 receptors of cholinergic neurons (Porkka-Heiskanen, 2013). Sustained activation of A1 receptors (the most abundant adenosine receptor subtype) throughout waking hours

induces the neuronal hyperpolarization necessary for the transition into NREM sleep, inhibition of calcium entrance to neurons and the reduction of acetylcholine release in basal forebrain (a neurochemical that promotes wakefulness) (Rodriguez et al., 2016). Moreover, pharmacological experiments have shown, both in humans and rats, that administration of adenosine receptor A1 agonists in the basal forebrain induces rapid sleepiness (Garcia-Rill et al., 2006). During the first hours of sleep, reconversion of endogenous adenosine back into ATP occurs predominantly in the same brain areas that were more metabolically active during wakefulness, thus, supporting the restorative function of sleep (Cartwright, 2001).

In this way, it has been proposed that SWA occurs as a consequence of accumulation of adenosine during wakefulness and the subsequent potentiation of potassium currents that hyperpolarize cortical neurons during the "off" states (Rodriguez et al., 2016). Additionally, pharmacological blockage of adenosine receptors in the basal forebrain significantly decreases SWA (especially in the first episode) and reduces total NREM sleep duration, but lengthens sleep latency (the amount of time to fall asleep, normally defined from "lights out" until stage 1 NREM sleep) (Landolt et al., 1995, Garcia-Rill et al., 2006). Finally, the world-wide consumption of caffeine, blocking A1 and A2 receptors activated by the adenosine molecule constitutes a practical proof of disruption of Process S by "hacking" brain systems involved in the build-up of sleep pressure (Landolt, 2012, Landolt, 2015, Clark and Landolt, 2017).

### 1.1.1. NREM sleep

"*Each night, when I go to sleep, I die. And the next morning, when I wake up, I am reborn.*" Mahatma Gandhi[1.2]

NREM sleep is characterized by low to high amplitude (proportional to sleep depth), low frequency (around 0.5-5 Hz) waves, reflecting the overall synchronization of large neuronal groups across the neocortex (Jones, 2016). At the neurochemical level, there is a great decrease of acetylcholine availability across the forebrain, hippocampus and brainstem, whereas monoaminergic activity (noradrenergic and

serotonergic) continues, but at an intermediate value between its maximum at waking and its depletion during REM sleep (Gais and Born, 2004). NREM sleep is divided in humans between “light sleep” (comprising stages 1 and 2) and “deep sleep” (or slow wave sleep, SWS; comprising stage 3) (Iber et al., 2007a). SWS sleep predominates during the first half of the night, yet it decays exponentially across cycles (Aeschbach et al., 1997). Reciprocally, sleep pressure builds up exponentially until an approximately asymptotic value throughout waking, to be then dissipated mostly by means of SWS during sleep (a payment of the homeostatic debt accumulated) (Aeschbach and Borbély, 1993). In addition, SWS amount and typical intensity levels diminish with increasing age, a fact that has been correlated to the decline in executive functions (e.g., attentional control) and memory consolidation among the elderly (Hobson, 2009).

Two prominent markers of stage 2 NREM sleep (which comprises 50% of total sleep time in healthy individuals) are the sleep spindles and K-complexes (Aeschbach and Borbély, 1993). Spindles refer to brief (approximately 1 s) synchronous bursts of oscillatory activity in the 12–16 Hz frequency range, which are generated in the reticular nucleus of the thalamus and then propagate to the neocortex (Hutchison and Rathore, 2015). A more detailed research distinguishes between slow (<12 Hz) and fast (12-16 Hz) spindles, with a greater preponderance of slow spindles in the frontal lobe, whereas fast spindles distribute along the parietal lobe and sensorimotor cortex (Mednick et al., 2013, Ladenbauer et al., 2016). Increased spindle density enhances neocortical long-term synaptic plasticity and correlates with performance after intense declarative memory training; this process is driven by calcium influx into cortical pyramidal cell dendrites (Jones, 2016). Remarkably, there is an inverse relationship between spindle activity and slow wave activity in NREM sleep (especially during the first sleep cycle), though to reflect a tradeoff incompatibility at the neuronal level (Aeschbach and Borbély, 1993).

The second characteristic waveform occurring during stage 2 NREM sleep, especially during the first sleep cycles, is the K-complex (Rasch and Born, 2013). A typical K-complex has two phases: an initial surface-negative wave (reflecting synchronous excitation of cortical

neurons) followed by a surface-positive wave (indicating neuronal hyperpolarization), exceeding 0.5 s in duration (Amzica and Steriade, 2002). Although K-complexes are mainly spontaneous events emerging for cortical dynamics, and hence endogenously driven, they can also be elicited in response to environmental external stimuli, possibly, in order to avoid awakening (provided that these stimuli are not meaningful for the subject, e.g., the crying of a baby for a mother) (Cash et al., 2009, Forget et al., 2011). Variations in membrane potential created by K-complexes induce slow oscillations (<1 Hz), although distinguished from typical delta waves by greater amplitude and more localized topographical distribution (Amzica and Steriade, 2002). Additionally, K-complexes have also been linked to declarative memory consolidation processes produced by memory reactivations and synaptic changes across widespread cortical circuits, but with a greater incidence on prefrontal structures (Born et al., 2006).

Regarding the function of NREM sleep, an extensive line of studies in animals and humans points to the critical role of NREM sleep for reactivation and consolidation of declarative memory (both semantic and episodic); furthermore, there are theories that go to the cellular level to explain these mechanisms (e.g., most prominently, the synaptic homeostasis hypothesis) (Gais and Born, 2004, Born et al., 2006, Stickgold, 2013, Tononi and Cirelli, 2014). Firstly, the lower acetylcholine availability observed during NREM sleep is supposed to disinhibit the transfer of information stored in the hippocampus into the neocortex, a process that is impaired with high cortisol levels (Lavenex and Amaral, 2000, Osborne et al., 2015). Secondly, sleep spindles occurring during stage 2 NREM sleep reflect thalamocortical interactions that improve synaptic plasticity and Hebbian learning related to declarative memory traces learned throughout the day (Mednick et al., 2013). Moreover, the duration of stage 2 NREM sleep is positively correlated with declarative memory consolidation (e.g., assessed by word-pair tests in humans), as evidenced by later performance testing (Hutchison and Rathore, 2015). Thirdly, the amplitude of SWS has also been found to be an indicator of information transferring between the cortex and the hippocampus, with the more declarative learning during waking (e.g., studying for an

exam), the greater SWS intensity during sleep (Gais and Born, 2004, Casey et al., 2016). Finally, SWS could even be beneficial for simple motor learning (whereas more complex motor learning profits from REM sleep), as shown by SWS selective deprivation (Rasch and Born, 2013).

### 1.1.2. REM sleep

"*We are such stuff as dreams are made on, and our little life is rounded with a sleep*." William Shakespeare[1.3]

REM sleep is a sleep phase common to mammals and birds (i.e., homeothermic animals) characterized by random eye movements, generalized muscle atonia and, in the case of humans, frequent report of vivid dreams upon awakening (around 80% of times, although dreaming activity also occurs during NREM sleep), although it is generally thought that animals exhibiting REM sleep may also dream, albeit in a very different way than humans (Siegel, 2005, Tsoukalas, 2012, Siclari et al., 2014). REM sleep is sometimes also referred as paradoxical sleep due to some shared biological aspects with wakefulness, most notably, low amplitude and fast frequency brain waves in the electroencephalogram, high level of brain activity (particularly in the visual cortex, motor cortex and limbic system) and other hormonal similarities (e.g., in the cortisol levels) (Jones, 2016). Despite some physiological commonalities with waking, a crucial difference is the significant depletion of monoamines during REM sleep (i.e., serotonin, norepinephrine and histamine), which explains the muscle atonia and, consequently, prevents the occurrence of motor activity out of dream content, which could be potentially damaging (as occurs in patients affected by REM sleep behavioral disorder) (Goder et al., 2011, Hutchison and Rathore, 2015). REM sleep is also characterized by the occurrence of ponto-geniculo-occipital (PGO) waves (Sakai et al., 1976) (traveling waves of synchronization highly coherent in the theta band, detected in humans suffering mental pathologies with intracranial EEG (Lim et al., 2007)), thought to be important for emotional memory consolidation and dreaming activity (Lim et al., 2007, Hobson and Friston, 2012). Another relevant feature of REM sleep is that its brain metabolism, as assessed by glucose consumption in positron emission tomography (PET) studies, may be

comparable or even exceed normal waking levels (Rasch and Born, 2013).

Several studies indicate that the pons (located in the brainstem) is both necessary and sufficient for the generation of REM sleep (Tsoukalas, 2012). Impairment or damage in the frontal lobe might suppress dreaming, but does not interfere with REM activity itself (Siegel, 2011). Thus, the activation-synthesis hypothesis presents a neurophysiological model aiming to explain the neurophysiological basis of REM sleep in the brainstem, as well as the cyclic alternation between REM and NREM sleep episodes (Hobson and McCarley, 2003). The theory proposes a kind of Lotka-Volterra model (differential equations), although translated to groups of different neuronal populations: the REM-on and REM-off neurons in the pons (playing the role of prey and predator in the classical Volterra equations, respectively). The REM-on neurons are cholinergic neurons that excite the REM-off neurons, which in turn, inhibit the REM-on neurons through the action of monoamines depleted during REM sleep (e.g., norepinephrine and serotonin). In this way, the dynamics emergent from the differential equation modeling predict the following consequences: an oscillatory behavior of both neuronal groups, a necessary initial state of low amines but high levels of acetylcholine in order to start a cyclical pattern in which high REM-on and REM-off activity alternate (Hobson and McCarley, 2003, Hobson, 2009). As an example, injection of carbachol (a cholinergic agonist) into the pontine brainstem produces within seconds induction of most REM sleep features: rapid eye movements, muscle atonia and PGO waves (Siegel, 2011).

Theories regarding the biological purpose of REM sleep still remain under development, but some important insights have appeared since its discovery (Aserinsky and Kleitman, 1953). Firstly, it seems that REM sleep plays a critical role for procedural and emotional memory consolidation (Smith, 2001, Rasch and Born, 2013, Hutchison and Rathore, 2015, Casey et al., 2016). It has been observed increased REM sleep after animals undergo training on an classical conditioning procedure (e.g., associating an odor with an electric shock), a motor

task or other forms of implicit learning (Born et al., 2006). Increased REM sleep also occurs when animals are immersed in an enriched environment, although the same holds for SWS (Rasch and Born, 2013). A second theory states that REM sleep evolved out of the tonic immobility reflex (a defense mechanism against predators based on death feigning, also known as animal hypnosis), as it is suggested by numerous biological similarities: low muscle tone, thermoregulatory cessation, increased coherence in the theta band (within the hippocampus, amygdala and cortex) and brainstem feedbacks with the forebrain (Tsoukalas, 2012).

Finally, another interesting theory proposed by Hobson suggests that REM sleep and dreams act as a state of protoconsciousness (Hobson, 2009, Hobson and Friston, 2012). Hobson distinguishes between primary and secondary forms of consciousness, the first (more primitive and shared with other mammals, associated with the temporal lobe among other structures) involves sensory perception and emotion, whereas the second (specific to humans and evolutionarily modern, associated with the frontal lobe) includes metacognitive aspects, abstract thinking and self-reflection, but is built upon the first (at the neurobiological level and also from an evolutionary perspective). Thus, REM sleep is a preparation, both at the ontogenetic and phylogenetic level, for the development of secondary consciousness (explaining why fetuses and babies exhibit such increased amount of REM sleep, even nearly 50% of total sleep time) (Klemm, 2011, Hobson and Friston, 2012). The theory proposes an evolutionary function for dreams (where primary consciousness predominates, with emotions and visual imagery most vividly during REM sleep, though dreaming is not specific to this state), as they provide a virtual reality simulator to better cope with real experiences, allegedly, where secondary consciousness (in the meantime, undergoing “maintenance” to optimize its later functioning during waking) would be advantageous (Valli et al., 2005, Hobson, 2009).

## 1.2. Biological functions and evolutionary theories of sleep

"*If sleep doesn't serve some vital function, it is the biggest mistake evolution ever made*." Allan Rechtschaffen[1.4]

Humans spend one third of their lifetime immersed in an activity which science has not yet elucidated what Aristotle called its "final cause" (teleological explanation of a natural phenomenon) (Barbera, 2008, Meisel et al., 2013). The existence of sleep seems paradoxical and difficult to predict from a Darwinian point of view, thus, posing an evolutionary puzzle (Savage and West, 2007, Schmidt, 2014). This is because in all natural niches, animals face the risk of being predated or harmed if not constantly alert within the environment that surrounds them, or at least could seize that period for other biological needs (mating, hunting, etc.); hence, unless they could be successful in hiding themselves every night, being "semi-unconscious" for long periods of time, it would be at first glance disadvantageous for survival (Siegel, 2009, Krueger et al., 2016). Even for predators, sleep provides an evolutionary benefit, since if predators were to hunt all day along, at some point they would also extinct out of starvation for the sharp dropping in prey population (Rechtschaffen, 1998, Siegel, 2005). In consequence, we can deduce that the benefits derived from sleep to an organism have to outweigh its potential costs in order to be selected by evolution, both at the individual level and also at the ecosystem level for an overall symbiotic balance (Lesku et al., 2009, Schmidt et al., 2016). Even when some species (such as birds, whales and dolphins) may cope with strategies that allow for a continuous monitoring of the environment by a cycling alternation of sleep between brain hemispheres (called unihemispheric sleep), sleep still is present, indicating that its biological function has to be indeed vital for life maintenance (Tobler, 1995, Rattenborg et al., 2000, Rattenborg et al., 2009).

Animal studies show that rats deprived of sleep die sooner than when they were deprived of food, due to a cascade of harmful secondary biological effects, maybe suggesting that the biological need for sleeping might be more fundamental than feeding (at least on rats), as its lacking creates more deleterious consequences (Everson et al.,

1989). Experiments with humans reveal the critical biological importance of sleep by exposing a series of disastrous health consequences due to chronic sleep deprivation, affecting both the body and the brain: a sharp decline in executive functions (working memory, fluid intelligence, etc.), increased risk of degenerative neuronal diseases (e.g., Parkinson and dementia), insulin resistance and increased risk of type II diabetes, obesity, hypertension and cardiac diseases, inflammation, etc. (Bobic et al., 2016, Chaput and Dutil, 2016). Moreover, an extensive longitudinal epidemiological study in humans, based on more than 1.3 million subjects followed up to 25 years, indicates that people who sleep for less than six hours each night were 12% more likely to die prematurely than those who sleep on the average range of 6-8 hours (Cappuccio et al., 2010). Nevertheless, it must also be mentioned that the same study found increased mortality risk for long duration of sleep, hence, displaying an U-shaped association between sleep and mortality (Cappuccio et al., 2010). Based on these reasons, we must conclude that sleep serves a crucial evolutionary and biological function.

Regarding theories about how old is sleep still remain preliminary, however, since both vertebrates and invertebrates manifest sleep habits (periods of body rest in which neuronal activity exhibits a characteristic pattern), sleep could have emerged from a common ancestor more than half a billion years ago, when multicellular organisms firstly appeared (there is no sign of restorative function that could be compared to sleep in unicellular organisms, only circadian rhythms) (Kavanau, 1997, Trinder et al., 2015). Furthermore, it is still unknown if mammals and birds evolved the neuronal basis for sleep (e.g., both exhibit slow wave sleep and REM sleep) from much older reptilian mammals or just as an example of convergent evolution (where the same trait appears independently in different species in the absence of causal connection or common inheritances), although the latter option reaches wider consensus (Rattenborg et al., 2009, Lesku et al., 2011).

Sleep may not be universal in all living creatures (e.g., sleep in bacteria is yet not established), however, is supposed to exist as an emergent property in any organism with a neuronal network (even in very small

networks as in *C. elegans*, with only 302 neurons) (Trinder et al., 2015, Krueger et al., 2016, Nichols et al., 2017). Among the most sound theories proposed to explain the function of sleep, which should be considered more in an eclectic way than as necessarily competing models (although along a hierarchy of importance), sleep has been said to serve the functions of: energy conservation and allocation; brain plasticity, memory consolidation and learning; and lastly, biological optimization through CNS "clearance" by the glymphatic system and immune system boosting (Rasch and Born, 2013, Xie et al., 2013, Schmidt, 2014). In the following, these theories are discussed in further detail.

### 1.2.1. Energy conservation and allocation theory

*"Think in the morning. Act in the noon. Eat in the evening. Sleep in the night."* William Blake[1.5]

According to the energy conservation model, sleep evolved for a thermodynamic reason, serving a similar function, in a first approximation, as hibernation in some endothermic animals (Berger and Phillips, 1995, Schmidt, 2014, Schmidt et al., 2016). Animals that hibernate (e.g., bears or bats) use this evolutionary strategy because when there are no available resources (e.g., due to food scarcity or hostile environmental conditions), entering into hibernation or torpor helps to save energy that could not be easily replenished under extant circumstances by means of lowering their basal metabolic rate (Kortner and Geiser, 2000, Toien et al., 2015). Hence, due to thermodynamic reasons, there had to be an evolutionary pressure for mammals to sleep, since in virtue of their relatively high surface to body mass ratio, energy exchange with the environment makes difficult to keep constant their body temperature, forcing animals to retreat from the environment and entering into a state of lower metabolic rate and resting (Joshi et al., 2016, Li et al., 2016). This theory also serves to explain, at least partially, why newborns and small mammals in general precise more time sleeping, as their surface to body mass ratio is greater (for human babies, which are born immature, a longer sleeping is also favored as it protects them from environmental

hazards), whereas big mammals (e.g., elephants) tend to have fewer hours of sleeping (Savage and West, 2007, Lesku et al., 2009).

Although it is considered to contain some degree of truth and validity, there are two main criticisms to the energy conservation model. Firstly, it seems difficult to reconcile with the existence of REM sleep, where core body temperature (controlled by the hypothalamus) increases as compared to NREM, and secondly, the actual energy savings derived from sleep are not that significant to be considered the fundamental explanation in the evolution of sleep (the reduction of metabolic rate during sleep compared to quite wakefulness is actually only of 5-15%) (Reynolds et al., 1991, Jung et al., 2011). In order to circumvent and refine the problems of the energy conservation theory, Schmidt introduced the energy allocation model (Schmidt, 2014, Schmidt et al., 2016). This theory explains the emergence of sleep in evolution as biological problem regarding the optimal distribution of energy resources needed for vital processes to maximize reproductive success. In this theory, the sleep-waking cycle results from a cyclical alternation between downregulation and upregulation of energy expenditure and biological self-investment (together forming a binary partition of all available energy), which are complementary variables playing a tradeoff relationship.

Thus, during wakefulness, energy expenditure predominates on tasks relative to niche exploitation (e.g., gathering energy resources, vigilance, mobility, mating, etc.), called waking effort. The rest of the energy not included in the waking effort is devoted to biological investment in the internal milieu of the organism (e.g., growth, cellular repairing and immune boosting, neuronal plasticity and reorganization, maintenance of reproductive organs, etc.), which is downregulated during waking, although still present. In this model, sleep exists as a way to optimize returns derived from allocation of energy investment, as it replenishes deficits accumulated during wakefulness. Consequently, during sleep, biological investment processes are upregulated whereas energy expenditure processes related to waking effort are downregulated, keeping a trade-off cycling relationship tight to Earth’s rotation leading to the 24-h light-dark cycle. Observe that in the energy allocation model, the benefits and evolutionary purpose for

sleeping concern both the brain and the rest of the body and organs in their optimal allocation of energetic resources. As an example, the liver has three times more changes in gene expression during sleep than the whole brain (Maret et al., 2008).

The allocation theory also proposes an evolutionary explanation for the development of REM sleep in particular. The evolutionary convergence for the appearance of REM sleep in both mammals and birds independently is causally correlated to their nature as endothermic animals. When compared to ectotherm animals of similar size, birds and mammals consume between 5 to 10 times more energy due the costs imposed by thermoregulation (Kortner and Geiser, 2000). Therefore, analogous thermoregulatory reasons could explain the previous case of convergent evolution for REM sleep. As it is known, thermoregulatory processes to keep a warm constant body temperature independently of the environmental temperature are suspended during REM sleep (Parmeggiani, 2007, Joshi et al., 2016). Since the total metabolic rate does not significantly vary when entering into REM sleep as compared to NREM sleep, it is therefore concluded that energy resources devoted to thermal regulation are allocated to other biological tasks. Given that thermoregulation, aiming to keep a constant internal temperature, is a very expensive energetic process, a kind of physiological process able to shunt energy devoted to thermoregulation to other ends pertaining biological investment would render great benefits. Hence, REM sleep is evolutionary advantageous as it allows a special opportunity time for the organism to dedicate more energy to biological investment goals.

Biological investment tasks that could be enhanced during REM sleep include: maintenance of reproductive organs (e.g., manifested in penile and clitoral erection) and motor systems (e.g., manifested by muscle atonia and intermittent twitches), cortical reorganization (e.g., manifested by the characteristic rapid eye movements and increased cortical activity), memory consolidation, particularly of emotional type (e.g., manifested by increase amygdala activity) and others still not fully understood (van der Helm et al., 2011, Tsoukalas, 2012, Jones, 2016). Furthermore, the alternation between NREM and REM sleep is

explained in this model as it guarantees that during the REM sleep period, thermal inertia of the organism does not deviate too much from its steady value, and that REM sleep is longer during the last hours of sleep (morning hours) as the difference temperature with respect to the environment begins to decrease due to Earth's rotation.

In summary, the evolution of sleep economizes energy for the organism on a daily basis (not coincidentally during a dark period when body resting serves an additional protective function), but more importantly, according to the allocation theory, serves to reallocate biological resources that represent complementary functions between self-investment (cellular and tissue repairing, learning and other CNS related functions pertaining neuronal reorganization, etc.) and energy expenditure (foraging exploitation and competition, vigilance, reproduction, etc.) (Siegel, 2009, Schmidt, 2014).

### 1.2.2. Glymphatic system and immune boosting

*"Sleep is that golden chain that ties health and our bodies together."*
Thomas Dekker[1.6]

A recent theory proposes that a main function of sleep is related to a kind of brain "clearance" of waste metabolites through perivascular tunnels via the glymphatic system (Xie et al., 2013). This system of perivascular tunnels through which neurotoxic metabolites are pumped for later elimination is structurally conformed by astroglial cells, but it is mainly disengaged from activity during wakefulness (Jessen et al., 2015). Hence, the sleep state enables an optimal internal environment for convective flow (composed of cerebrospinal fluid or CSF and interstitial fluid, which mutually interchange) from the brain to the circulatory system; this flow drags with it soluble proteins, toxins and other waste products derived from waking away from the brain (Krueger et al., 2016). Indeed, clearance during sleep facilitated by the glymphatic system allows a two-fold activity increase compared to waking state (Mendelsohn and Larrick, 2013). This theory also explains why after sleep deprivation, executive functions (such as sustained attention, working memory and problem-solving) are significantly impaired, as a direct consequence of a "dirty"

environment at the molecular level in the whole brain, and more particularly in the frontal lobe (Aguirre, 2016).

A proposed physiological mechanism regarding the glymphatic "clearance" is briefly summarized in the following. During wakefulness, continuous neuronal activity resulting from thoughts and other mental experiences require the conversion of large molecules (e.g., glycogen) to small molecules (e.g., lactate, $H^+$ and $CO_2$) in order to sustain it, thus, causing a concomitant raising of osmotic pressure and interstitial fluid volume (Jessen et al., 2015). Additionally, neuronal activity demands oxygen and nutrients, increasing cerebral blood flow, which in turns represents another contribution for increasing local extracellular volume (Krueger et al., 2016). Consequently, given that water is basically not compressible and brain volume is constant (not along decades, but on a daily scale), neurons need sleep as a time and special milieu to cope with these changes that gradually deviate from homeostatic conditions, by changing the ratios of cellular to extracellular volume (Xie et al., 2013). As it is known from basic hydrodynamics, a reduction of cellular volume in neurons allows better clearance by entry of CSF and exchange of interstitial fluid, as it diminishes the resistance to fluid circulation with increasing extracellular cross-sections among the neuronal networks and related parenchyma (Mendelsohn and Larrick, 2013).

The protective effect of sleep in the delay of neurodegenerative diseases can also be accounted for, at least partially (i.e., ruling out genetic and life-style contributions), by the clearance model (Krueger et al., 2016). Animal studies demonstrate that old compared to young mice show an approximate 80-90% efficiency decrease in glymphatic clearance (possibly due to stiffening of arterial walls and reduced arterial pulsatility with aging), as evidenced by radiolabeling with CSF tracers (Jessen et al., 2015). Thus, in humans, it has been proposed that gradual dysfunction of the glymphatic system with aging is causally connected with deterioration of brain tissues, grey matter loss and pathogenesis of neurodegenerative diseases among the elderly (Mendelsohn and Larrick, 2013). For example, the extracellular removal of beta-amyloid plaques by the glymphatic system is a crucial

and necessary factor for staving off the progression of Alzheimer disease and other cognitive dementias, as clearance avoids a steady accumulation of misfolded proteins in brain tissue (Mendelsohn and Larrick, 2013, Adam et al., 2016). It has also been suggested the importance of glymphatic activity for proper recovery in traumatic brain injury, as fluid pumping during sleep intervenes in the fast elimination of cytosolic proteins released after severe neuronal traumatism aiding to later scarring by the glia (Plog et al., 2015).

Another proposed function for sleep supported by a multitude of studies is related to maintenance and bosting of the immune system (Irwin, 2015, Krueger et al., 2016). There is plenty of evidence both from animal and human studies that insufficient sleep debilitates immune defenses, rendering the organism more vulnerable to immune attack, diseases, infections and slower healing from wounds (Dissel et al., 2015, Irwin and Opp, 2017). As an example, sleep deprivation raises the risk for getting the flu or cold as a consequence of a significant reduction of T-cells accompanied by a parallel increase of inflammatory cytokines (Balachandran, 2011, Aguirre, 2016). Moreover, sleep facilitates optimal internal conditions for fighting an illness (as it has been recommended by physicians since antiquity, e.g., as shown by the raising of body temperature in a fever when patients are asleep (Lange et al., 2011)), or even promotes the healing of the brain after suffering traumatic brain injury (Edlow and Lammers, 2016). Research by the medical community during the last decades suggests that lacking of the recommended approximately 8 hours of sleep or poor sleep efficiency heightens the risk for developing cancer, cardiac diseases, diabetes and brain-mental diseases (such as depression, Alzheimer and Parkinson) (Irwin et al., 2011, Irwin, 2015).

In analogy to the CNS, the immune system also benefits from sleep for better “learning and memory”, and both share the three fundamental steps in memory formation of encoding, consolidation and retrieval (Rasch and Born, 2013). An example of the importance of sleep for immune learning has been shown with vaccines, as sleep potentiates the effects of vaccination (especially during SWS), whereas sleep deprivation generates less antibodies to save a record of the antigen and the immunization process takes longer to occur (Lange et al.,

2011). The interaction effect between immunity and sleep is also evidenced by the sleep enhancing effects produced by endogenous immune response modifiers (e.g., the interleukin 1 family) and bacterial products (e.g., muramyl peptides, stimulators of SWS) (Brown, 1997). Finally, the decrease of cortisol levels during SWS is supposed to facilitate the trafficking and accumulation of T cells in the lymph nodes and has an influence on cytokine release (Rasch and Born, 2013).

Overall, there is an undeniable amount of empirical evidence corroborating the cross effect interactions between sleep and the immune system, however, it is thought that the described immune boosting is an opportunistic function appearing later in evolution, but not a primordial reason for why sleep evolved (e.g., does not explain why sleep occurs accompanied by a state of semi-unconsciousness) (Krueger et al., 2016).

### 1.2.3. Brain plasticity, memory and learning

*"What could not be repeated at first is readily put together on the following day; and the very time which is generally thought to cause forgetfulness is found to strengthen the memory."* Quintilian[1.7]

A third theory that has gathered good empirical validation over the decades states that a principal function for the existence of sleep is related to neuronal network plasticity (i.e., learning of any nature) and memory consolidation (Smith, 1996, Stickgold and Walker, 2005a, Rasch and Born, 2013, Hodor et al., 2014). As it is evidenced both from animal and human studies, after training on a certain motor or cognitive task (e.g., procedural or declarative memory, problem solving, etc.), sleep, either during a whole night or simply after a nap, significantly improves learning as compared to a similar period of wakefulness (Stickgold, 2005, Stickgold and Walker, 2005b). On the other hand, sleep deprivation may also hinder memory consolidation; for example, SWS deprivation is detrimental for explicit learning of visuospatial content (Casey et al., 2016). Another proposed underlying mechanism refers to the importance of sleep for synaptic plasticity, either by strengthening of synapses activated during learning activities

(e.g., playing a musical instrument or a dance movement) or by a homeostatic synaptic downscaling during sleep after waking synaptic potentiation (Tononi and Cirelli, 2006, Cirelli, 2013, Cirelli and Tononi, 2015). In addition to structural neuronal plasticity, neurogenesis occurring in the dentate gyrus of the hippocampus could also be particularly boosted during sleep, a biological process that is significantly deteriorated under a sleep deprivation protocol (Tung et al., 2005, Meerlo et al., 2009).

After establishing the general importance of sleep for learning and memory consolidation in a causal way, the neuroscientific community aimed to discover the specific roles played by NREM and REM sleep pertaining different kinds of long-term memory aspects (Smith, 2001, Stickgold and Walker, 2005b, van der Helm et al., 2011). Contrary to the conception in which it is used in the common language, memory is not a unitary system but a rich constellation of different mental processes; as a consequence, long-term memory is usually divided into explicit or declarative (further subdivided into episodic and semantic; consciously accessible and shareable) and implicit (further subdivided into procedural and emotional; unconscious and not shareable) (Rasch and Born, 2013, Casey et al., 2016). As it might be imagined, the neuroanatomical substrates underpinning different memory aspects are not the same, thus, motivating its separate study and their particular correlations to the different sleep phases (Stickgold and Walker, 2005b, Stickgold, 2013).

In the case of procedural memories, both motor (e.g., learning a gymnastic exercise) and cognitive (e.g., the procedure to solve a geometry problem) types precise REM sleep for appropriate learning and behavioral efficiency (emotional memories regulated by the amygdala have also been linked to REM sleep), whereas learning of declarative memories are predominantly influenced by NREM sleep (Smith, 2001, Meerlo et al., 2009, Rasch and Born, 2013). Among the brain structures associated with each memory aspect, for implicit memories, cognitive procedural memory include the striatum and the prefrontal cortex, cognitive motor memory involve the primary motor cortex, supplementary motor area and cerebellum, and emotional memory relates principally to the amygdala (e.g., important for

aversive conditioning learning), whereas for explicit memories, the hippocampus and surrounding structures (entorhinal cortex, parahippocampal gyrus and thalamus) assume an essential role (Smith, 2001, Born et al., 2006). In addition, experimental evidence provided by studies in rats, but probably also transferable to humans, suggests that the previously learned task is replayed during sleep in the same neuronal circuits used during waking (e.g., including an interplay between the hippocampus and the cortex during SWS) in order to be fully consolidated and enhanced (Eckert and Tatsuno, 2015, Farthouat and Peigneux, 2015).

Pharmacological studies have also proved useful to further clarify the causal relationship between sleep and memory consolidation (Mednick et al., 2013, Rasch and Born, 2013). In this regard, administration of antidepressant drugs (such as SSRIs, MAOIs, tricyclic antidepressants and others), which are known to significantly suppress the amount of REM sleep, do not cause later impairment in executive functions or declarative memory consolidation and recollection, but might disrupt the learning of procedural and emotional memories (Smith, 2001, Goder et al., 2011). Drugs may not only impair memory consolidation, but also enhance it; in particular, it has been shown that administering zolpidem (a GABA agonist hypnotic) following a verbal memory task (i.e., an explicit memory task) and an immediate nap, significantly increases spindle density and performance on subsequent testing, allegedly by causal influence on spindles for semantic memory (Mednick et al., 2013). The same study also did not find any causal influence on spindle enhancing (whose frequency and amplitude was not altered) for procedural motor memory, but in turn, showed a worsening of perceptual memory (as assessed by a texture discrimination task) by a concomitant shortening of REM sleep produced by the drug.

A deeper investigation of declarative memory learning in humans has revealed that SWS is the most relevant and beneficial sleep stage for its long-term consolidation and the progressive reduction of SWS with age is connected to the memorization and retrieval deficits observed in the elderly (Gais and Born, 2004, Ladenbauer et al., 2016).

Endogenous neurochemical agents critically involved in declarative memory consolidation include the acetylcholine neurotransmitter and cortisol hormone (as well as other glucocorticoids). The acetylcholine decline associated with NREM sleep might be necessary for better transfer of signal output from the hippocampus and entorhinal cortex to the neocortex, as well as inside the hippocampus for feedback communication between the CA3 and CA1 regions (thought to be related to memory replaying) (Gais and Born, 2004). However, during wakefulness, acetylcholine actually improves declarative memory encoding, either directly or indirectly, as this neurotransmitter is relevant for attention (Sarter et al., 2003). On the other hand, the endocrine hypothalamic-pituitary-adrenocortical axis (HPA) is crucially involved in memory encoding by the action of cortisol released from the adrenal glands, which facilitates memory formation (e.g., an animal succeeding to escape from a dangerous place or after experiencing a stressful situation in humans), but at the same time, hinders memory retrieval (Osborne et al., 2015). Cortisol levels, which follow a circadian cycle, diminish until a minimum during the first hours of sleep, precisely when slow wave activity is at its maximum, which in turn inhibits the HPA axis, an inverse relationship that creates an optimal environment for information transferring from the hippocampus to the cortex for long-term storage (Bierwolf et al., 1997). This causal relationship is evidenced by experimental manipulation of cortisol levels in human studies, as infusion of cortisol during SWS significantly impairs declarative memory consolidation by hindering hippocampal-neocortical communication (Lavenex and Amaral, 2000, Gais and Born, 2004).

## 1.3. Sleep onset

### 1.3.1. Sleep onset transition: neurophysiological aspects

*"There is a time for many words, and there is also a time for sleep."* Homer[1.8]

In healthy humans, the sleep onset (SO) transition is carried through NREM sleep, as a direct entrance to REM sleep is considered to be an indicator of brain pathology (e.g., as it is typically seen in narcolepsy (Ogilvie, 2001); babies are an exception as their brains are yet immature) (Cartwright, 2001). Scientists have struggled for many decades attempting to provide a single index or biomarker able to pinpoint the exact moment of SO, however, efforts in this direction are most probably misleading (Wright et al., 1995). Currently, the neuroscientific community agrees that a sort of process that unfolds gradually (rather than crossing a "gap" at an instantaneous moment) seems closer to the truth and better able to provide a good descriptive model for this biological phenomenon (Magnin et al., 2010, Prerau et al., 2014). In this regard, the process of falling asleep has been compared to the difficulty in defining the moment of death, as different criterions (depending on which physiological variable is temporally tracked) often lead to very diverse conclusions, which may differ considerably (Davis et al., 1938, Ferrara and De Gennaro, 2011, Marzano et al., 2013).

The most straightforward and simple definition for SO (hence, the first to appear historically) relies on a set of behavioral characteristics (Ogilvie, 2001). Aristotle was the among the first thinkers to present a naturalistic theory of sleep based on empirical observations; in his book *De somno et vigilia*, sleep is described as a natural state characterized by reduced physiological functions and absence of motion (due to "excess of waking" –similar to the modern Process S-, as no animal is "able to exercise continuously") and even advances that growth is promoted during sleep (Barbera, 2008). In the modern era, researchers have catalogued a series of features that accompany the SO process at the behavioral level in a majority of species: postural recumbence on a secure place, body quiescence, closed eyes and changes involving the slowing down of a number of important

physiological processes (Tobler, 1995, Siegel, 2005). For instance, starting in stage 1 NREM sleep, the respiratory system rapidly modifies its normal rhythmic activity, undergoing a substantial ventilation dropping as a result of increase airway resistance induced by low tone in the breathing muscles (controlled by the brainstem when breathing is not consciously driven, as normally occurs) (Prerau et al., 2014). Related to the respiratory system, cardiovascular activity is also reduced during the SO transition: heart rate decreases gradually with deeper sleep stages (however, this correlation is only observed in NREM sleep; during REM sleep, heart rate might exceed waking levels (Faes et al., 2014)), and entrance into stage 1 NREM sleep parallels a decrease in oxygenated hemoglobin (Ogilvie, 2001, Faes et al., 2014).

A common physiological technique used to aid the tracking of the SO process relies on the measurements provided by electromyography (EMG) (Prerau et al., 2014). EMG changes recorded in the respiratory muscles seem to give the best results, as a gradual decrease in their muscle tone correlates nicely with the transition from waking to stage 1 NREM sleep, but facial or extremity measurements can also be employed (Ogilvie, 2001). However, EMG has the important caveat that, being the person relaxed just before SO (as can be normally assumed, besides insomnia and anxiety disorders), the corresponding EMG recordings might be indistinguishable from the post SO period, thus, motivating the introduction of more reliable and precise techniques that address brain activity directly (Cartwright, 2001, de Munck et al., 2007). In this way, the electroencephalogram (EEG) reveals that seconds to minutes after the reduction of muscle activity (the SO transition is also characterized by a parallel slowing of eye movements), alpha activity (8-12 Hz, associated to quiet wakefulness and relaxation) progressively shifts towards a state of low voltage and mixed-frequency activity, particularly in the occipital cortex (Wright et al., 1995, Magosso et al., 2007). At the subjective level, during this period of transition from wakefulness to stage 1 NREM sleep, conscious activity gradually disengages from external stimuli to be mostly pervaded by internal thoughts (e.g., autobiographical memories), probably reflecting increased activity of the default mode network (DMN) (de Munck et al., 2007, Speth and Speth, 2016).

Once fully into stage 1 NREM sleep, there is a slow rolling of the eyes (not to be confused with eye saccades during REM sleep) and theta activity predominates, although occasional briefs bursts of alpha activity and waking arousals may still occur (Iber et al., 2007b). Experiments eliciting evoked response potentials (ERP) with auditory stimuli during this sleep stage confirm a significant drop in environmental responsiveness, although inwardly oriented mental activity still continues (Larson-Prior et al., 2011). Additional research with functional magnetic resonance imaging (fMRI) revealed that this reduced reaction to external stimuli was paralleled by increased DMN activity and correlates to a rising of frontal, temporal and parietal BOLD (blood-oxygen-level-dependent) activity at the beginning of stage 1 NREM sleep, but towards the end was mostly located in posterior brain regions (Picchioni et al., 2008). Moreover, relative to waking, there is a general decreasing of thalamic and fronto-parietal BOLD activity during stage 1 NREM sleep (a finding also corroborated by positron emission tomography (PET) imaging (Smith, 2001)), that could account for the "disconnection" with the external world as well as the progressively poor performance on a simple motor task (e.g., repetitive finger tapping) (Larson-Prior et al., 2011). Finally, it is worth mentioning that about 90% of people aroused from stage 1 NREM sleep report that they had not been sleeping, despite continuous EEG monitoring confirming all previously described neurophysiological changes (Cartwright, 2001).

Following the SO transition, after stage 1 NREM sleep, the brain enters into stage 2, when sleep becomes deeper and, as a result, previous neuronal changes are intensified (Iber et al., 2007b). Reduced reaction to external stimuli and perceptual awareness of the surroundings becomes even more evident than for stage 1 NREM sleep, as assessed by fMRI or ERP testing (Ogilvie, 2001). During this stage, different brain areas are observed to fall asleep at different rates, reflecting varying levels of sleep pressure built up during the day (Magnin et al., 2010, Marzano et al., 2013). In this way, changes in chemical neuromodulation involving aminergic and cholinergic systems lead to the fast hypoactivation of the frontal lobe, particularly the ventromedial and dorsolateral prefrontal cortices (Hobson and

McCarley, 2003, Speth and Speth, 2016). In spite of all these processes, a complete disengagement from the external world does not happen; this can be shown by an attenuated, but not absent, olfactory response to odorants (e.g., peppermint), or auditory stimuli, provided to be meaningful to the subject (e.g., participants hearing their own name show increased activation of bilateral orbitofrontal cortex) (Larson-Prior et al., 2011). In contrast to stage 1, during stage 2 sleep the percentage of participants reporting they had not being sleeping when awakened by the researchers lowers to 40%, that is, a majority agrees with the EEG manifestations of cortical activity indicative of sleep (Cartwright, 2001).

### 1.3.2. Sleep onset transition: insights from latest quantitative research

*"Why do we do basic research? To learn about ourselves."* Walter Gilbert[1.9]

Further progress in the understanding of the SO transition from a rigorous quantitative analysis was provided by the studies of De Gennaro et al. (De Gennaro et al., 2000a, De Gennaro et al., 2004). Their main interest was to compare the most prominent changes of mathematical parameters indicating the strength of correlation between frontal and occipital regions before and after SO, as well as the dominant directionality in the underlying mutual coupling. Thus, the coupling of the anterior-posterior axis (referring to the connections between the frontal and occipital lobes) is statistically compared between the pre-sleep onset period (interval of 5 minutes preceding the first K-complex or spindle) and the post-sleep onset period (interval of 5 minutes following the pre-sleep interval) across different frequency bands. The EEG pairs of channels used to evaluate coherence were Fz-Pz and Fz-Oz (that is, a fronto-parietal and fronto-occipital connection, respectively), whose results were then averaged.

Firstly, the synchronization level (as assessed by spectral coherence) in the delta band up to the alpha band (from 1 to 11 Hz) was significantly different between the pre-sleep interval and the one post-sleep onset, but changes occurred in opposite directions depending on the frequency range, whereas there was no significant difference for higher

frequency bands (sigma and beta). Thus, in the delta-theta band (1 to 7 Hz), there was a significant reduction ($P$=0.03; from pre to post SO) in mean coherence levels, however, for the alpha band (8-11 Hz), there was a significant increase ($P$=0.05) in coherence after SO. Secondly, coherence was averaged for pre and post SO intervals for each individual pair, yielding systematically greater coherence in the fronto-parietal than the fronto-occipital connections across the examined frequency bands ($P$<0.01 for 1 up to 30 Hz, 1 Hz bin resolution; $P$ values were lowest for delta-theta bands). Moreover, sorted from highest to lowest coherence values and averaged between pre and post sleep onset, the frequency bands of greater coherence were: beta (highest), sigma, delta-theta and alpha (lowest). Thirdly, the problem of directionality was addressed by a linear technique called Directed Transfer Function (DTF, more information in chapter 5), which has in common with coherence a continuous distribution in the frequency domain and, furthermore, is able to discriminate the right order of causal synchronization (e.g., from Fz to Pz or from Pz to Fz, unanswered by a correlation index). Results provided by the DTF were split between before and after SO. The analyses reveal that parieto-occipital to frontal information flow predominates in the period preceding SO in delta-theta and alpha bands, whereas this directionality pattern was inverted after SO, with frontal to parieto-occipital information flow for all frequency bands. Finally, an average of all frequency bands indicates no significant difference in directionality for the fronto-parietal connection before SO (however, there was a net flow from occipital to frontal regions), whereas after SO, the frontal cortex drives parieto-occipital regions in all bands ($P$<0.001 for all tests).

Taking a different approach, SO transition was also investigated by Massimini et al. by means of transcranial magnetic stimulation (TMS) combined with high-density EEG (HD-EEG) (Massimini et al., 2005). The main hypothesis was that the fading of consciousness accompanying SO is due to a breakdown of cortical effective connectivity. Effective connectivity refers to causal influences established when one neuronal group drives the dynamic properties (firing rate, phase, etc.) of other neuronal groups in the brain, thus, not

being limited just to characterize symmetrical correlations (Barrett et al., 2012). Furthermore, it is hypothesized that consciousness does not simply depend on synchronization of different neuronal groups at certain frequency bands (which continues during sleep, but shifted to slower frequencies), sensory input or firing rate levels, but on the ability or efficiency to integrate information using cortico-cortical and cortico-thalamic loops.

In order to test this hypothesis, TMS stimulation was repeatedly applied during quiet wakefulness and then compared to the cortical response generated in NREM sleep (Massimini et al., 2005). TMS offers the advantage that of being a direct way of probing the brain, since when compared to sensory stimulation, it does not activate the reticular formation (a network of nuclei in the brainstem linked to arousal and consciousness) and bypasses the thalamus, hence, creating minimum interference and peripheral effects (Esser et al., 2009). Thus, the experimental protocol is based on TMS stimulation on a single brain area that is fixed under both conditions (the premotor area, an area with extensive cortical connections), and then recording with HD-EEG measurements how the spreading of electrical activity affects different cortical areas along its course of propagation. Results show that for the waking condition (eyes closed, relaxed state before falling asleep), TMS pulses produced a weak initial disturbance before the perturbation decayed to zero (in about 300 milliseconds), but showing a radius of propagation of several centimeters that could travel unremitting to the other hemisphere. Later, when entering into stage 1 NREM sleep (in the SO transition), the initial perturbation had a bit higher amplitude and created a wave that could still propagate inter-hemispherically, nonetheless, this new wave remained for less time in a sustained way before vanishing (indicating decreased cortical effective connectivity). In particular, the rippling triggered during stage 1 in the right premotor could reach the contralateral region within tens of milliseconds, but evincing a hindered capability for self-sustainability (indeed, prefrontal and parietal areas were not affected) (Massimini et al., 2005).

The maximal breakdown of effective connectivity was detected, in agreement with the model, during deep NREM sleep (SWS). Here,

TMS induced a significant loss of transcallosal connectivity (i.e., the propagation could not extend inter-hemispherically through the corpus callosum) and long-range connections (i.e., waves had a much smaller radius of propagation than in waking or stage 1 sleep, essentially restricted to the spot of stimulation). Moreover, the initial cortical response to the pulse was stronger than in the waking or stage 1, but also faded significantly faster and was more stereotypical when the stimulation site was moved across the cortex (Massimini et al., 2005). Hence, in the context of this theory, the frequent dizziness and transient "blank state" observed after subjects are awakened from SWS was attributed to a much reduced effective connectivity in the thalamo-cortical system than during wakefulness. Finally, to account for the breakdown of effective connectivity during SO, the authors proposed the progressive reduction of the firing of sleep related modulatory systems and a dynamical bi-stability of cortical networks (related to transitions from up to down states, mediated by outward potassium currents in neurons).

A major issue for the unraveling of the SO process concerns the precise mechanism that induces the cortical deactivation required to fall asleep. This question was addressed by Magnin et al. with the use of intracranial recordings (both intrathalamic and intracortical recordings monitored in parallel) performed in epileptic patients (Magnin et al., 2010). During waking, thalamic and cortical structures are time-locked; this functional connection (where the connection between thalamus and posterior cingulate cortex may be especially important) is necessary, although possibly not sufficient, to produce and maintain conscious awareness (Vogt and Laureys, 2005). In addition, during stage 2 of NREM sleep and REM sleep, thalamic and cortical activity cyclically alternate periods of tight coupling and decoupling (at the cellular and functional level), suggesting a more complex interplay than was once previously assumed (Rey et al., 2007).

In order to tackle the problem of how to quantify whether a particular brain region is either awake or asleep, Magnin et al. (Magnin et al., 2010) introduced a non-linear method called dimension of activation

(DA). In a simplified conceptualization, DA is a degree of "complexity" in a signal, quantified by the number of correlations between its spectral frequency components. For example, SWS has only a major component (located in the delta-theta range), thus, DA assumes a low value, whereas waking is composed by a mix of many frequencies, rendering higher values of DA. Therefore, when falling asleep, DA is supposed to gradually decrease, indicating that a brain region has begun to sleep after crossing a certain threshold to be considered significant. Using this approach, it could be shown that the thalamus (together with the brainstem, hypothalamus and other structures involved in the ascending reticular system) undergoes a deactivation process that precedes cortical deactivation up to several minutes. Consequently, cortical structures are able to remain "awake" for several minutes after thalamic activity has already started to decrease to sleep levels. A detailed analysis revealed that in 92.8% of cortical areas investigated (126 in total), the deactivation process happened slower than for the thalamus, introducing several delays depending on the cortical region, but of an average of 8 to 9 min (ranging from 15 s even up to 27 min after deactivation of the thalamus). Moreover, for deep SWS, DA lowered until a stable value faster in all examined cortical areas (i.e., reaching a perfect 100%) than in thalamus, with an average delay of approximately 15 min.

In this way, SO can be formally differentiated from other processes of conscious fading (in spite of similar behavioral manifestations at the external level), most notably pharmacologically induced anesthesia, where the reverse pattern occurs, that is, the cortex deactivates before thalamic and related circuits (Mashour, 2011, Barrett et al., 2012). Remarkably, even when comparing the same cortical region, considerable variability among different subjects in the time delay between thalamic and cortical deactivations emerges. A second source of variability (controlling for the same subjects) is created when varying the considered brain region across the cortical mantle, that is, there exists a wide distribution of time delays depending on specific cortical areas. A large amount of empirical evidence supports the hypothesis of heterogeneous locality during the SO process, as opposed to a global sudden deactivation (Nir et al., 2011, Larson-Prior et al., 2011). Some biologically relevant examples can be provided in

this respect. Firstly, hypnagogic hallucinations are explained by a still locally activated cortex (including the occipital cortex for visual imagery and the precuneus for DMN related autobiographical memories); secondly, sleep spindles are mostly localized in fronto-parietal regions; thirdly, the fronto-parietal gradient in power spectral density (PSD); and finally, the frequent overestimation for the time needed to fall asleep (given that many regions have already deactivated, but transient traces of consciousness may continue, typically in stage 1 of NREM sleep) (De Gennaro et al., 2004, Marzano et al., 2013, Sarasso et al., 2014). In light of these findings, the authors concluded that cortical disengagement allowed by previous thalamic deactivation is a necessary prerequisite for the fading of consciousness accompanying SO. Finally, for the process of awakening (requiring the reactivation of the same circuits that shut down during SO), the method of DA showed that thalamic and cortical structures activate synchronously (i.e., a functional coupling is established simultaneously) (Magnin et al., 2010).

A study performed by Sarasso et al. has shed light on this still not well-understood question regarding the relationship between the occurrence of sleep spindles and the SO transition (Sarasso et al., 2012, Sarasso et al., 2014). In particular, this group investigated hippocampal spindles preceding cortical SO in human subjects. SO was operationally defined as the first identifiable epoch of stage 2 (characterized by the emergence of a spindle or K-complex) at the Fz-Cz scalp derivation. Spindle activity was calculated as the average power spectral density in the range of 11-16 Hz both for hippocampal and cortical recordings, allowing spindle detection in the frequency domain when a certain power threshold (based on the standard deviation of power distribution) was crossed. A second classificatory condition was that putative spindle events had to show durations between 0.5-2 s in order to be properly considered spindles.

Simultaneous recordings in neurosurgical patients by means of combined EEG and stereo-EEG (SEEG, used for intracerebral recordings) of cortical and hippocampal activity (respectively) revealed that spindles were already present in the hippocampus several minutes

prior to the SO transition (mean delay of 11.3±2.4 min) (Sarasso et al., 2014). In addition, hippocampal spindles systematically preceded cortical spindles (mean delay of 4.5±1.6 min). Statistical analyses were split between main features before and after SO. Before SO, spindle density (i.e., number of spindles per minute) was significantly greater in the hippocampus compared to all other cortical regions; in particular, no spindles were detected in the temporal cortex preceding SO (although they could be observed afterwards). This trend continued after SO, except for the parietal cortex, which showed comparable spindle density to the hippocampus. On the other hand, average frequency of hippocampal spindles (both preceding and following the SO transition) was also significantly faster than for cortical spindles. Furthermore, spectral analyses corroborated, in consistency with previous research, that frontal spindles oscillate at a slower rate than centro-parietal spindles (Aeschbach and Borbély, 1993, Nir et al., 2011). Nevertheless, regarding amplitude (or power), there was no significant difference between hippocampal and cortical spindles, preserving its stereotypical shape.

A subsequent topographical analysis of the time lapses between hippocampal and neocortical spindle activity indicated the existence of an antero-posterior axis (Sarasso et al., 2014). Hippocampal spindle bursts were consistently followed after a brief period of a couple of minutes by either frontal or parietal spindles. In addition, mean delays were shorter for frontal and nearby areas (insular cortex) than for centro-parietal regions (Sarasso et al., 2014). This frontal preponderance in minimal delays along with its significantly faster deactivation process as compared to the rest of neocortical areas could be partly explained by higher homeostatic pressure that has built up in these regions during waking (Finelli et al., 2001a, De Gennaro et al., 2004, Marzano et al., 2013). In general, these findings support the thesis of a heterogenic coexistence of dissociated brain patterns during SO, in particular, concerning the hippocampus (a key region for the development of declarative and episodic memories) (Eckert and Tatsuno, 2015). Furthermore, the earlier dissociation of the hippocampal formation from the neocortex could explain the common subjective experience of amnesia preceding SO, called “mesograde amnesia”, i.e., being unable to evoke a posteriori the previous moments

just before falling asleep as well as the parallel fading of consciousness (Ogilvie, 2001, Hobson and Friston, 2012). Finally, hippocampal spindle activity was correlated to anterior thalamic activity and, moreover, hippocampal spindles occurred phase-locked to thalamic slow waves. Hence, it is argued that the anterior thalamic nucleus serves as the driving synchronization source for the hippocampus (in this regard, a structural connection between both structures could be established via the fornix and cingulum white matter bundles) (Cartwright, 2001, Andrillon et al., 2011).

A crucial aspect for the understanding of SO transition involves the evolution of the main properties of slow waves and spindles; this topic was addressed in depth by Siclari et al. (Siclari et al., 2014). The experimental protocol consisted of multiple awakenings of participants (repeated in periods of 15 to 30 min) in order to better capture the gist of how SO unfolds, taking advantage of HD-EEG recordings (256 channels). In this way, the data sample could be multiplied several times for subsequent analyses (a total of 141 recordings were finally analyzed with only 6 participants), therefore, allowing a more precise statistics. SO was not defined as a single moment, but as the time interval between two hallmark events: firstly, the disappearance of alpha activity and subsequent emergence of stage 1 (beginning of SO), and lastly, the first emergence of two or more consecutive slow waves in the same derivation (ending of SO). Using this definition, the temporal period analyzed covers on average duration of approximately 10 min, constituting the x-axis for the temporal tracking of all studied variables (split into 10 epochs). Finally, EEG data at the scalp level was processed through an inverse solution method (LORETA software for the necessary transformation matrix, then combined with GeoSource tool for spatial representation) to obtain the underlying neuronal currents produced at the source level.

Firstly, the study (Siclari et al., 2014) examined the evolution of general characteristics of cortical slow waves, including temporal density (i.e., total number of slow waves detected per minute), amplitude, slope and number of negative peaks. Temporal density revealed an increasing exponential evolution, rising slowly during the

first five min, but rapidly during the last five min. The steepness associated to the exponential proliferation of slow waves throughout the cortex was most prominent in the prefrontal cortex and less evident on parieto-occipital regions, hence, exhibiting an antero-posterior gradient. Amplitude (or, because of the proportionality, power) showed an initial increase that peaked around 3 min after beginning of SO and then slowly decreased, being anti-correlated at the same time with the growth of temporal density of slow waves. Slope evolution was positively correlated (in fact, directly proportional) to amplitude evolution; again, frontal lobe areas displayed the highest values. Number of negative peaks experienced an initial decrease that reached a minimum around 3 min after beginning of SO and afterwards raised monotonically. In consequence, amplitude and number of negative peaks were anti-correlated variables, suggesting shared underlying properties from which these relationships emerge.

Secondly, a more detailed analysis of slow waves at the source level could dissociate between two distinct processes (furthermore, characterized by opposite properties, e.g., low versus high density), which were named synchronization processes I and II (Siclari et al., 2014). Synchronization process I (leading to type I slow waves) is dependent on a “bottom-up” (i.e., subcortico-cortical) synchronization mechanism, closely related at the anatomical level to the ascending reticular activating system (ARAS). The ARAS is tightly associated to arousal promoting structures, involving predominantly the locus coeruleus (the greatest source of noradrenergic innervation in 0the brain). In turn, the locus coeruleus sends numerous projections to the prefrontal cortex, which is the same brain region where slow waves are firstly observed (that is, early during the SO transition). Hence, a common input to large frontal areas originated from fixed subcortical structures might explain the experimentally observed near simultaneous synchronization of widespread cortical regions and their main oscillatory characteristics. In this way, type I slow waves (resulting from the exposed process I, occurring at the beginning of SO) show large amplitude, steep slope, long duration, few peaks and involve extended frontal regions (first to fall asleep). The emergence of type I slow waves is also linked to a sharp decrease in gamma power

that helps to consolidate the early phase of the SO transition (Dalal et al., 2010, Valderrama et al., 2012).

On the other hand, synchronization process II (leading to type II slow waves) is dependent on a "horizontal" (i.e., cortico-cortical) synchronization mechanism, appearing in a later phase than type I slow waves along the course of SO (Siclari et al., 2014). The proposed mechanism of synchronization for type II slow waves is more complex than for type I; instead of being based on a major subcortical "synchronizer" directing in parallel large cortical groups, process II is based on solutions of the wave equation propagating on a 2-D space (i.e., the cortical surface). Although still not well-understood, this phenomenon is attributed to slow oscillations most probably initiated on layer V of the cortex, which then are able to propagate by "recruiting" nearby cells, thus, creating waves of synchronization along a radial path on multiple cortical sites (Nir et al., 2011, Siclari et al., 2014). Process II is less efficient to induce widespread groups of synchronized neurons than process I, as its ability to propagate is limited by the number of neurons that can be successfully depolarized by "synchronizer" neurons (acting as the source). As a result, multiple focuses of synchronization are generated throughout the cortex (without a prevalence of any region) in order to avoid the fading of type II waves, allowing both sustainability and temporal continuity for the propagation of these waves. Given that they are based on a completely different induction mechanism, type II slow waves are described by opposite properties than type I waves: small amplitude, flat slope, short duration and multiple cortical origins (i.e., there is no privileged source focus). In addition, type II slow waves exhibit greater spatial density than type I slow waves, i.e., are more numerous across the neocortex. This fact fits with the general trend of increasing power in the delta band during SO: given that type II waves show smaller amplitude than type I and appear later, a major increase is needed to compensate power levels, hence, the density variable undergoes such rise that even enhances previous power levels (Marzano et al., 2013).

This paper also investigated, although in less detail than for slow waves, the overall evolution of spindle activity and related

characteristics along the SO transition (Siclari et al., 2014). Temporal density analyses revealed that global scalp involvement in spindle activity (defined as the percentage of EEG channels showing simultaneous spindle synchronization) followed an inverted U-shape course across the 10 min interval examined. This finding is congruent with previous literature (Aeschbach et al., 1997, Nir et al., 2011). Consequently, few cortical areas (mostly on the frontal lobe) showed initial spindle involvement, but it gradually increased until covering most fronto-parietal areas, reaching a maximum of activity that lasted for around 4 min, which then slowly faded away during the final minutes. A filtered distinction between fast and slow spindles could establish that during the first half of the interval, fast frequency spindles (of low spatial density and mostly local) preponderated, whereas slow spindles were sparse, this trend was inverted during the second half interval (when spindles were spatially diffused, involving more distributed areas). Finally, maps for source localization indicated that fast spindles originated on the precuneus and posterior cingulate cortex, whereas slow frontal spindles originated on the anterior cingulate cortex, medial and dorsolateral prefrontal cortices (Del Felice et al., 2014, Siclari et al., 2014).

Changes in default mode network (DMN) functional connectivity development paralleling the SO transition were investigated by Sämann et al. (Sämann et al., 2011). DMN is among the most deeply studied functional networks in neuroscience; it has drawn increasing attention over the last years in part given its close relationship with psychiatric diseases (e.g., depression or schizophrenia) (Horovitz et al., 2009, Salone et al., 2016). DMN originates as a result of spontaneous mental activity (mind-wandering, rumination, autobiographical memories, etc.) and concomitant synchronization of multiple neuronal groups distributed both across cortical and subcortical structures (Larson-Prior et al., 2011). Some regions that are strongly linked to DMN activity include the posterior cingulate cortex (PCC), precuneus, medial prefrontal cortex (mPFC), bilateral inferior parietal lobule (IPL) and the hippocampal formation (Buckner et al., 2008, Sämann et al., 2010). Given that DMN activation (which correlates to heightened alpha power) (Benedek et al., 2014) reflects a state in which subjects are not externally oriented (i.e., paying low attention to external cues),

but inwardly centered, it is reasonable to investigate its relationship with SO, a process entailing clear similarities (e.g., pronounced, although not absolute, environmental disengagement) (Ogilvie, 2001, Larson-Prior et al., 2011).

It has previously been shown that during stages 1 and 2, DMN activity continues to be present (as assessed by PCC seed correlations), nonetheless, exhibiting a significant decoupling in functional connectivity compared to wakefulness, probably as a result of progressive reduction in consciousness at the subjective level (Picchioni et al., 2008, Horovitz et al., 2009, Larson-Prior et al., 2011). Integrity in structural (i.e., white matter tracts) and functional (i.e., correlative) connections between PCC and other brain areas seems essential to maintain consciousness (especially regarding connections with prefrontal areas), however, most probably only as a necessary, but not sufficient, condition (Vogt and Laureys, 2005). Empirical findings supporting this statement include a significant decrease in PCC seed correlations under anesthesia and vegetative states, and the fact that PCC glucose consumption is the highest in the entire brain (Mashour, 2011, Amico et al., 2014).

Using simultaneous EEG/fMRI recordings, DMN evolution was continuously monitored, firstly, during wakefulness preceding SO, and afterwards from light sleep until SWS. Statistical analyses showed that the null hypothesis of no significant changes in DMN state along the SO transition has to be rejected in a double way: by changes in focal configuration (i.e., different arrangement of the principal nodes that together create the network) and synchronization strength (i.e., the degree of coupling between nodes or inter-node). During the transition from wakefulness to stage 1, the DMN network remained relatively unaltered in its focal configuration; however, there was a significant decrease in PCC and precuneus synchronization to all other DMN nodes, even when PCC continued to be the most important hub of synchronization of the network (Vogt and Laureys, 2005). Increased sleep pressure also affected PCC seed connectivity by weakening the connections between this major node and the rest of DMN nodes,

therefore, constituting another biomarker of the homeostatic Process S (Horovitz et al., 2009, Sämann et al., 2010).

Entering into stage 2, mPFC stepwise decoupled from the network and, as a result of significant fragmentation, produced the disruption of the overall DMN configuration (Sämann et al., 2011); this effect is congruent with frontal lobe areas displaying maximal homeostatic pressure, thus, facilitating the development of the SO process (Marzano et al., 2013). Moreover, the greatest disengagement of the mPFC node was observed during SWS, that is, when conscious vigilance is most suppressed. Consequently, it is argued that functional integration of mPFC in the DMN network (in particular, to posterior nodes) is critically necessary for maintenance of consciousness, and moreover, could even play a role in dream mentation, besides wakefulness (Sämann et al., 2011, Amico et al., 2014). In parallel to the decoupling of anterior areas, posterior areas showed increased short-range connectivity during stage 2 sleep (e.g., connections between midcingulate cortex, precuneus and bilateral inferior parietal lobules) (Larson-Prior et al., 2011). Additionally, the parahippocampal gyrus (a part of the hippocampal formation) disengaged gradually from the DMN network during stage 1 sleep and remained practically decoupled during stage 2 sleep, a finding that correlates with the phenomenon of “mesograde amnesia” (inability to recall the immediate moments preceding SO) (Sarasso et al., 2012). Finally, increased sleep depth was also related with progressive decoupling of fronto-parietal (or antero-posterior) nodes of the DMN network, interpreted as a state of impaired information transmission through long-range connections. This result is in agreement with previous studies by Massimini et al. regarding the breakdown of effective connectivity during the SO transition (Massimini et al., 2005).

# 2. Brain connectivity and EEG source localization

## 2.1. Brain connectivity: introductory overview

*"Invisible threads are the strongest ties."* Friedrich Nietzsche[2.1]

There are two main descriptive views for brain connectivity, addressing the topic in different, albeit complementary, ways (Bernasconi and König, 1999, Stam et al., 2016). The first outlook transforms the connectivity problem into an abstract and simplified graph of entangled and coupled nodes of a network, whereas the second is more rooted onto well-established biophysical properties of individual neurons, neuronal networks and large brain circuits (Friston, 1994, Friston, 2011). Therefore, the first approach relies on mathematical modeling, and consequently, is universal in its potential range of applicability, that is, the same modeling principles are directly transferable to other scientific realms besides neuroscience (including physics, systems biology, economy and stock market, etc.) (Penny et al., 2009, Ashley and Tsang, 2014). The second approach, which is built upon neurobiological measurements and empirically validated concepts about how biological neuronal networks "truly" behave, presents the advantage of specificity for the neurosciences, yet still lacks a unified theoretical framework regarding how structural and functional connections exactly relate with each other (Park and Friston, 2013, Dipasquale and Cercignani, 2016). Recent avenues of research aiming to merge at the same time both approaches, such as the Human Brain Project, Human Connectome Project or the BRAIN initiative, are expected to show in the coming decades a tremendous potential for major scientific breakthroughs, unraveling brain functioning and its underlying connectivity patterns either in the healthy or maladaptive brain (Markram, 2012, Alivisatos et al., 2015, Glasser et al., 2016). Both approaches are summarized in more detail in the following.

The theoretical modeling approach is mostly based, as a starting point, on a branch of mathematics called graph theory, which considers the collection of interacting neuronal elements selected for analysis as interconnected nodes arranged in a complex (yet very idealized) network, whose topological characteristics are to be described (Sporns et al., 2007, Faes et al., 2010). A major difficult arising in graph theory

is that the complexity of the network (proportional to the total number of possible connections) increases exponentially with the number of nodes, turning the analyses quite challenging both at the theoretical and computational levels (nonetheless, continuous improvements are being made in this regard, as accuracy of mathematical modelling and computational capabilities advance) (He and Evans, 2010). In general, this approach does not depend on brain invasive experimental methodologies, but on noninvasive indirect measures of neuronal activity (such as EEG, MEG and fMRI), which constitute the basis for subsequent modelling and analysis (Picchioni et al., 2008).

Given that anatomical structural connectivity pathways are mostly static (at least, when considering small time scales; e.g., years of practice will alter the motor cortices of pianists), task-related brain activity is attributed to modifications in functional connectivity (either geometrical reconfiguration or coupling changes) between neuronal elements, corresponding to the "edges" (i.e., all interconnecting nodes) of the network (Baccala and Sameshima, 2001, He and Evans, 2010). Therefore, brain dynamics is understood within this approach as temporal variability on the coupling strength of the edges of the network, node reorganization towards different network patterns or the simultaneous combination of both means of change, eventually leading to the phenomenon of neuroplasticity (e.g., pianist example cited above) (Babiloni et al., 2005, Shine et al., 2015). Additionally, the overall architecture of biological neuronal networks remains plastic after neuroplastic adaptations resulting from experience (setting aside limitations imposed by the aging process, rigid lifestyles or genetic factors), hence, assuring efficiency and flexibility in brain responses to diverse potential environmental demands, especially when they entail novelty (Erla et al., 2009, Larson-Prior et al., 2011, Edlow and Lammers, 2016).

In contrast, the biophysical approach takes into account explicit measured information regarding the way complex brain networks actually connect at the synaptic and neuronal (involving electrochemical transmission and synaptic plasticity), circuital (studying neuronal circuits from the microscopic to the mesoscopic

scale) and regional level (considering how distinct macroscopic regions interact, e.g., the thalamus and the visual cortex via the optic radiation) (Sporns et al., 2007, Cachope, 2012). Recent technological advances of the scientific tools with which the brain can be probed (hugely improving accuracy and spatial or temporal resolution to levels difficult to imagine just decades ago) are gradually allowing the recollection of massive amounts of neurobiological "big data", moreover, frequently adding the advantage of using non-invasive imaging techniques for *in vivo* visualization (Korzeniewska et al., 2008, Khadem and Hossein-Zadeh, 2014). Hence, the introduction of these new probing technologies (e.g., enabling the 3-D morphological reconstruction of individual synapsis and local microcircuits much beyond millimeter resolution) has motivated the progressive mapping of the inner intricate anatomical structure of the mammalian brain along with all its connectivity pathways, ultimately aiming toward a complete connectivity atlas of the human brain (Dipasquale and Cercignani, 2016, Glasser et al., 2016).

As a consequence, this approach has vastly enriched our knowledge of the microscopic cytoarchitecture of neurons and dendritic trees, the spatial configuration of cellular circuits (whose plasticity is currently possible to track in real time) and the tractography of long-range fiber bundles, as well as other neurobiological properties that are crucial for describing the complex connectivity organization of the brain (Friston, 2011, Markram, 2012). In particular, the field of connectomics aims to create a complete map of synaptic connections through piece-by-piece combination of all compiled biophysical data, which eventually (perhaps at some point of this century) could be useful for the computational simulation of brain activity (possibly as an emergent phenomenon of the connections) or the customized treatment of mental diseases (e.g., schizophrenia) (Glasser et al., 2016, Houck et al., 2017, Kristan, 2017).

Finally, connectivity can be investigated in several different ways, which are described in the next subsections. As a preliminary summary, structural connectivity characterizes white matter bundles linking distinct brain regions, functional connectivity detects their correlational dependencies (which can be either direct or indirect) and

effective connectivity extends even more the scope of research, first, through systematic pruning of indirect connections, second, explicitly addressing causality (at least, in a mathematical sense) and finally, introducing directionality aspects (Sameshima et al., 1998, Vicente et al., 2011, Kralemann et al., 2014).

### 2.1.1. Structural connectivity

Structural connectivity reflects long-range fiber tracts linking a set of neuronal circuits, which can be experimentally inferred from magnetic resonance imaging (MRI) or, alternatively, diffusion tensor imaging (DTI) (Stam et al., 2016). At the anatomical level, structural connections are formed by white matter tracts, whose projections allow feedback communication between cortico-subcortical or cortico-cortical regions (e.g., the cingulum, connecting the cingulate gyrus and the entorhinal cortex, whose deterioration is related to mild cognitive impairment) (Kocsis et al., 1999). The importance of structural connectivity is manifested on the fact that it can successfully predict (yet not completely) functional connectivity patterns appearing during resting states (as it is revealed when conclusions drawn from structural connectivity are checked to the actual application of functional connectivity methods) (He and Evans, 2010, Friston, 2011, Reimann et al., 2017).

In a simplified description, it can be stated that structural connectivity (whose overall architecture is determined both by genetic inheritance and experience) shapes large-scale neuronal dynamics at the most fundamental level, which are later “refined” by functional and effective connectivity (more dependent on environmental experience or cognitive demands), although the interaction between both is ultimately bidirectional (Sporns et al., 2007, Jäncke, 2016, Sporns and Betzel, 2016). Nonetheless, context specific task engagement (e.g., reasoning about a logic puzzle) may create a progressive divergence between structural and functional networks, when during learning it is no longer possible to reasonably predict connections of the latter out of the architecture of the former (at least, in the short-term) (Sporns and Betzel, 2016). Structural connectivity properties can be experimentally

revealed to some extent through a normalized parameter called fractional anisotropy (FA), which constitutes a measure of white matter integrity (i.e., degradation of myelin sheets, as in multiple sclerosis, renders low FA values) and helps in the estimation (although is insufficient) of the strength of structural connections (Andrews-Hanna et al., 2010, Huang et al., 2015).

In order to ensure both robustness and flexibility, structural connectivity unfolds through normal human development in such a way that finds a compromise between wiring costs (e.g., by increasing myelination amount or long-range connections) and computational efficiency (e.g., by pruning excessive synaptic connections throughout adolescence) (Liu and Aviyente, 2012). The optimal balance imposed by these constraints results, as an emergent phenomenon, in the exhibition of small-world properties in neuronal networks, as evinced by structural connectivity topology (Lee et al., 2012, Park and Friston, 2013). Small-world organization implies high local neighborhood clustering and short average path length between any pair of node, assuring efficient global information transferring while reducing wiring costs (Langer et al., 2012). A second relevant property related to structural connectivity concerns what has been called “modularity” (He and Evans, 2010, Sporns and Betzel, 2016). High modularity implies a dense intrinsic short-range connectivity within a brain module (e.g., primary visual cortex), although extrinsic connectivity between different brain modules might be weak and sparse (e.g., prefrontal disengagement occurring in mild cognitive impairment) (Chan et al., 2013, Douw et al., 2014). Currently, the prevalent view in structural connectivity studies is that brain modules remain integrated through long-range axonal connections, which combined with complex patterns of synchronous activity, are amenable to form diverse important large networks, such as the default mode network (DMN), salience network (SN) or dorsal attention network (DAN) (Park and Friston, 2013, Salone et al., 2016).

Noteworthy caveats of structural connectivity (at least with the currently available techniques), are that it cannot distinguish between indirect and direct connections in a causal sense, pinpoint the

predominant direction of information transmission or discriminate between excitatory and inhibitory connections (Park and Friston, 2013). Additionally, structural connectivity based on MRI might also be deficient to detect the weaker long-range axonal connections, which nevertheless may play an important role in global connectivity (e.g., to better characterize brain pathologies) (Salone et al., 2016, Stam et al., 2016).

### 2.1.2. Functional connectivity

Functional connectivity aims to reveal the existence of statistical correlations appearing between nodes or defined regions of interest (ROIs) of neuronal networks, although it does not precise putative dominant directions of synchronization neither hypothesizes causal origins between remote neurophysiological events (Palus, 1997, Stam et al., 2016). Signals whose functional correlations are most studied in the neurosciences include BOLD signals (obtained from clusters of voxels in fMRI that act as ROIs) and EEG/MEG signals (with the respective analyses of groups of synchronized channels at the scalp level) (Friston, 2011). Classical tools typically employed to study functional connectivity include, firstly, coherence and phase-locking index in the frequency domain, and secondly, cross-correlation measures (e.g., zero-lag and partial correlations) in the temporal domain (Le Van Quyen et al., 2001, Kaminski et al., 2001, Bastos and Schoffelen, 2015). As an example, an interesting application of functional connectivity is the prediction of personality traits (e.g., neuroticism), intelligence levels, or even mental disorders, simply based on resting-state measures of functional connectivity (particularly, the DMN), whose network patterns can cluster together similar cognitive profiles (i.e., extract connectivity hallmarks) into different types of groups (e.g., low, medium or high presence of a trait) (Langer et al., 2012, Douw et al., 2014, Ikeda et al., 2015).

Structural connectivity is closely related to functional connectivity, constituting to a great extent the anatomical basis for the establishment of correlative relationships (e.g., as shown in interregional phase-couplings) (Chan et al., 2013, Huang et al., 2015). Nonetheless,

functional connectivity might exist even in the absence of structural connectivity, through indirect processes (e.g., mediated by a common linking driver) (Bernasconi and König, 1999, Stam et al., 2016). Functional networks (i.e., the topology formed by a collection of synchronized nodes) are constrained, although not limited, by the architecture of structural networks (Goldenberg and Galvan, 2015). In this way, even when structural connectivity, just on itself, does not completely determine functional connectivity, it serves to create a general connectivity "skeleton", and furthermore, it posits important restrictions to the potential range of functional connections that can be established within the network (Stam et al., 2016). In turn, at the next level of analysis, functional connectivity usually serves as the working basis for effective connectivity, and in spite of providing plausible clues for information transmission throughout neuronal networks, effective connectivity cannot be derived from functional connectivity (Friston, 2011).

Therefore, a crucial conceptual point regarding functional connectivity is that solely based on its existence (as indicated by suitable mathematical tools and even high significant values), it is not possible to conclude the presence of anatomical interconnected nodes at the neurobiological level (Babiloni et al., 2005, Korzeniewska et al., 2008). For instance, functional connectivity can exist between two nodes that are not directly connected by any synaptic or axonal path through common input provided by a third node, which acts as a "synchronizer driver" and bridges previously disconnected nodes (Astolfi et al., 2007, Blinowska, 2011). Additionally, functional connectivity between a pair of nodes may not be an accurate measure of the coupling between the corresponding brain areas, because in complex networks, indirect paths also make an additive contribution that affects the final results, hence, connectivity indices would actually be revealing the sum of both direct and indirect paths of synchronization (Schelter et al., 2009, Faes et al., 2010). An example where this distinction is appropriately taken into account is the method of partial coherence, a modified version of the classical coherence formula aiming to better represent the functional coupling between two given nodes that removes spurious (or indirect) synchronizations (Baccala and Sameshima, 2001, Astolfi et al., 2007). Thus, partial

coherence and similar methods (e.g., partial mutual information and direct directed transfer function) yield lower values than standard functional connectivity indices after correcting undesired contributions produced by indirect paths, hence, being more suitable and precise for connectivity studies (Blinowska, 2011, Chan et al., 2013).

Some caveats appearing in the use of functional connectivity, besides its theoretical limitation of explicitly neglecting causal relationships and directionality issues, are worth mentioning (Park and Friston, 2013). Firstly, functional connectivity is generally sensitive to the kind of method used, i.e., depending on how functional coupling is measured or defined, applied coefficients could vary significantly, simply because they reflect distinct properties (e.g., amplitude correlations versus phase correlations, or temporal versus spectral correlative characteristics) (Le Van Quyen et al., 2001, Ramanand et al., 2010a). Secondly, temporal windows chosen by the researcher are expected to have an impact on the yielded final results, with prominent fluctuations for short temporal intervals (i.e., standard deviation of the coefficients might be very large) and non-significant values for long temporal intervals (as every temporal correlation, even when highly significant, vanishes if averaged over an extensive period) (Friston, 1994, Jeong et al., 2001). Finally, confounding variables masking putative network connections calculated based on functional connectivity methods might render the whole analyses sterile and wrong (as the existence of confounding variables is generally not taken into account explicitly by these methods), although this caveat is significantly corrected when introducing additional effective connectivity methods (Faes et al., 2012, Sommerlade et al., 2012).

### 2.1.3. Effective connectivity

Effective connectivity measures the influence that one neuronal group or ROI exerts over another (or even on a whole collection of other groups at the same time, when multivariate analyses are performed), addressing specifically the directionality and the causal nature (either direct or indirect) of the connections (Vicente et al., 2011, Barrett et al., 2012). As a general statement, inferring effective connectivity requires successive steps of model specification (comprising

hypothesis about network arrangement), model identification (selection of the most suitable hypothesis) and, finally, causal modelling (explicit results yielded by the most likely network architecture) (Barrett et al., 2012). Similarly to functional connectivity (where structural connectivity delineates the general repertoire of possible functional connections), effective connectivity is also influenced to some extent by structural connectivity, although this time, structural information is not sufficient to provide significant clues about effective connectivity and directionality (even having a perfect knowledge of both structural and functional connectivity) (Stam et al., 2016). Additionally, as with other connectivity forms, effective connectivity is subject to plastic changes occurring across the lifespan; e.g., it has been observed that network organization in the infant is different compared to the adult brain (the latter being arranged in a more centralized way when directing causal synchronization) (Goldenberg and Galvan, 2015).

Effective connectivity can be studied either with tools specifically built for assessing neuronal responses in task-evoked experimental procedures (e.g., responses in the auditory cortex to music and how it drives limbic structures for emotional significance) or, conversely, in a more generalized and autonomous manner (i.e., context-independent, however, more dismissive of environmental inputs, not explicitly modeled) (Barrett et al., 2012, Liu and Aviyente, 2012). For the first approach, which includes external inputs to the brain directly into the equations, the most commonly used and a well-established method is the Direct Causal Modelling (DCM), originally developed for fMRI but applicable for EEG data too (Friston, 1994, Penny et al., 2009). A common alternative to DCM involves the use of Structural Equation Modelling (SEM), which rests on modelling effects of controlled experimental manipulations to the system and also uses network graphs to discover active links and their causal properties (Sommerlade et al., 2011, Valdes-Sosa et al., 2011).

Briefly, DCM relies on Bayesian modelling to estimate how changes in neuronal activity in one node are caused by activity of another one, providing parameters for connection strength (which are sensitive to the directionality of transmission) and brain modulation by experimental inputs (Friston, 2011, Friston et al., 2014). The use of

Bayesian statistics incorporated into the DCM theory makes it possible to find the causal network graph with highest evidence (i.e., explanatory power for observed functional connectivity at the causal level, as assessed by the Bayes theorem), and in case that several different networks were a priori compatible with the data, selects by default the most simple network (Park and Friston, 2013). Finally, it has been shown that DCM outperforms most alternative methods when dealing with networks with cyclic feedback loops (e.g., closed circuits allowing bi-directionality) and, in addition, DCM can take into account the previous temporal history of the network to calculate "memory footprints" (i.e., long-term effects of the past which still exert a causal influence) (i.e., long-term past effects which still exert a causal influence) (Valdes-Sosa et al., 2011).

On the other hand, effective connectivity is also studied through a second general family of models which are not explicitly conditioned by experimental set-ups and external causal manipulations, but purely descriptive of internally driven variations of neuronal activity in brain networks (Schelter et al., 2009, Vicente et al., 2011, Barrett et al., 2012). Therefore, whereas the previous approach takes a view on causality more focused on the modelling of biophysical influences (i.e., by changing causes –e.g., external inputs-, the effects on the network will also change), this approach understands causality as temporal precedence (causes always precede effects) (Valdes-Sosa et al., 2011). The most prominent example for this approach is Granger causality (GC), along with all the subsequent improvements (e.g., frequency representations, multivariate analyses, etc.) (Granger, 1969, Kaminski et al., 2001). It is worth pointing out that practically all "temporal precedence" methods aiming to be totally general can be historically traced to GC as a kind of "common ancestor", including partial directed coherence (PDC), directed transfer function (DTF) or isolated effective coherence (iCOH) (Baccala and Sameshima, 2001). The gist concept in this family of methods is that there is "causality" (in a mathematical sense) between variables A and B (where A causes B) whenever uncertainty or prediction error in the modeling of variable B is reduced by inclusion of previous states of variable A (Faes et al., 2010, Faes et al., 2012).

Finally, even if effective connectivity is a very powerful method for connectivity studies, as it was essentially designed to overcome most limitations of functional connectivity techniques, it also exhibits several caveats (Kralemann et al., 2014). First, a majority of effective connectivity methods are constrained to operate just at the level of measured signals (with the important exception of DCM, aiming to provide statistical tools to overcome it), hence, positing a hard problem when trying to model the underlying behavior at the neuronal level, which cannot be directly observed or probed with non-invasive techniques (Penny et al., 2009). It is likely to suppose that this limitation will be gradually improved with the introduction of more refined models enabling an accurate description of the biophysical link between neuronal activity and the resulting BOLD or EEG signals (called forward models), probably by addition of non-linear terms in the mathematical equations (i.e., broadening the simplicity of linearity) (Valdes-Sosa et al., 2011, Friston et al., 2014). Secondly, effective connectivity results might be misleading due to superposition of electrical fields produced by different brain regions on nearby electrodes and volume conduction effects (whose amplitude of propagation decreases with the inverse of the distance) (Vicente et al., 2011, Huang et al., 2015). Hence, conclusions drawn from effective connectivity analyses are prone to error when limited to the scalp level (Pascual-Marqui et al., 2014a). In this way, even when effective connectivity methods might detect correctly the coupling between regions, directionality aspects might be wrong (especially for complex networks), although there are methods more adept to overcome these limitations, which outperform many others (e.g., iCOH as compared to GC) (Kaminski and Blinowska, 2014, Goldenberg and Galvan, 2015). Finally, it is expected that effective connectivity models will be updated in two interacting ways: first, by incorporating new discoveries on structural connectivity (hence, refining our theoretical comprehension), and secondly, by using more realistic computational modeling for both resting state and task-induced activity (however, this limitation is continuously being improved) (Park and Friston, 2013, Park et al., 2015).

## 2.2. Source localization with focus on low resolution electromagnetic tomography (LORETA)

*"eLORETA is a genuine inverse solution and not merely a linear imaging method. We show that it has exact, zero error localization in the presence of measurement and structured biological noise."* Roberto D. Pascual-Marqui[2.2]

As it is known, both conscious and unconscious mental activity induce continuous electromagnetic signals as a result of underlying neuronal dynamics and metabolic processes in the brain, either during task engagement, resting state or sleep (Hobson and McCarley, 2003, Horovitz et al., 2009). Thus, uncovering brain areas implicated in the production of different mental states requires the introduction of techniques able to locate the most likely brain sources that are consistently active during the experimental task (Vanoosterom, 1991, Jatoi et al., 2014). The systematic search (i.e., by following a mathematical algorithm) for the spatial localization of putative active sources inside the brain that are responsible (or at least play a moderating role) for the generation of EEG patterns recorded at the scalp level, is called brain source localization (Anderer et al., 2001, Silfverhuth et al., 2012). Source estimation based on EEG is theoretically grounded on the inverse problem of electromagnetism, which consists on calculating the distribution of current sources (the unknown variables) that produce a certain electromagnetic field, out of a finite sample of measurements (the known variables), and which can be proven to possess no unique solution (indeed, an infinite number, even if the number of sensors were infinite) (Pascual-Marqui, 2002).

The motivation to find source localization methods adapted to EEG data stems from the limitations of commonly used experimental techniques for functional imaging and localization of brain activity, such as fMRI or PET imaging (Pascual-Marqui et al., 1994, Nichols and Holmes, 2002). Both fMRI and PET constitute three-dimensional brain tomographies exhibiting excellent spatial resolution, however, their temporal resolution is quite limited (in the range of seconds in the case of fMRI) (Farthouat and Peigneux, 2015, Houck et al., 2017). Therefore, these tomographies are not able to track the typical speed at

which most neuronal processes occur (Astolfi et al., 2007). For instance, in the case of fMRI, it can be shown that the temporal resolution of the BOLD signal is equivalent to a low-pass filter of ongoing neuronal activity, that is, only captures the lowest frequencies, while ignoring faster components in the frequency spectrum (Pascual-Marqui et al., 2002). On the other hand, EEG has a different (although somehow related) biophysical grounding, making it more adept for creating a high-temporal resolution tomography (Pascual-Marqui et al., 2011).

Briefly, small local clusters of synchronized cortical cells (whose typical dimensions are comprised between 40 to 200 $mm^2$) superpose their electrical fields, and the spatial summation resulting from the overall activity (post synaptic potentials) at a given moment constitutes the signal registered by the EEG sensors (Cartwright, 2001). Hence, electrical signals recorded by EEG correspond to activity originating from cortical pyramidal cells (importantly, these neurons are oriented perpendicularly to the cortical surface), and can achieve a temporal resolution of milliseconds, thus, offering a great scientific potential for creating a suitable brain tomography (Jones, 2016). Based on the fact that pyramidal neurons producing the recorded potential fields are orthogonally oriented to the surface, modelled dipoles used for source reconstruction are constrained to follow the same orientation, thus, resulting in a modest simplification to the inverse solution problem (nonetheless, the strength of the dipoles remains unaffected by this geometrical constraint) (Pascual-Marqui et al., 2011). However, the achieved accuracy obtained through source localization reconstruction based on EEG is compromised by a number of limiting factors (Vanoosterom, 1991). The most important to mention are: volume conduction effects (making more ambiguous the position of the generators due to electrical propagation, possibly picked by several nearby sensors), EEG noise (either of instrumental or biological origin, as both contribute) and head-modelling errors (arising when indicating the spatial coordinates of the sensors, necessary for the calculation of the source generators) (Grech et al., 2008).

In general, source localization techniques can be divided into two main categories: parametric and non-parametric methods (Faes et al., 2012).

Parametric methods, aim to estimate the dipole parameters (including position) of a given number of dipoles (e.g., using a non-linear least square approach), whereas non-parametric methods calculate dipole magnitude and strength for a set of fixed positions distributed across brain volume (e.g., by iteration and regularization) (Grech et al., 2008). Given that modeled equations appearing in parametric methods are non-linear, non-parametric methods are much preferred due to their linear character, simplifying the inverse problem both at the theoretical and computational level (Faes et al., 2012). It can be shown mathematically that the linearity of non-parametric methods is a consequence of the huge simplification entailed by imposing from the beginning all dipole locations (usually forming three-dimensional spatially regular grids), hence, these degrees of freedom do not have to be estimated by any algorithm (Pascual-Marqui et al., 2002). Among non-parametric methods, we can mention: LORETA (low resolution electromagnetic tomography, (Pascual-Marqui et al., 1994)), WMN (weighted minimum norm, (Jihene et al., 2013)), SLF (shrinking LORETA focus, (Wu et al., 2005)), Backus-Gilbert (Wan et al., 2008) and S-MAP (sequential maximum a posteriori expectation maximization, (Paavolainen et al., 2014)); these non-parametric methods were computationally compared against a validated toy model (of previously known solution) by Grech et al., rendering LORETA as the method of choice, both in terms of accuracy (especially for deep sources) and noise robustness (Grech et al., 2008).

In the context of brain source localization based on EEG/MEG, Low Resolution Electromagnetic Tomography (LORETA, last version called eLORETA, for “exact”) is a computational tool built to solve the inverse problem in a biologically realistic way and estimate intracerebral current source densities (CSD) located in brain gray matter, thus, yielding three-dimensional CSD maps (Pascual-Marqui et al., 2002, Pascual-Marqui et al., 2011). CSD is a measure of energy density proportional to the total influx of electromagnetic current crossing an orthogonal plane of unit surface (thus, measured in units of $\mu A/mm^2$), such as the sum of currents derived from the oscillation of electrical dipoles in cortical pyramidal neurons in a given local brain region (Babiloni et al., 2005, Astolfi et al., 2007). On the other hand,

three-dimensional CSD maps are based on a grid of several thousand volume elements or voxels (there are 6239 voxels of 5 millimeter resolution in LORETA), which are geometrically arranged according to an average adult brain neuroanatomical model, encompassing all Brodmann areas (following the digitized Talairach Human Brain Atlas of the Montreal Neurologic Institute) (Pascual-Marqui et al., 2002). Solution space for CSD maps includes cortical and hippocampal gray matter, using a probability atlas (provided by the MNI brain atlas) to classify each voxel as belonging to either gray matter, white matter or cerebrospinal fluid (the latter being excluded from calculations, as there are no underlying neuronal oscillations associated with these voxels, e.g., in the ventricular system) (Lantz et al., 1997). The solution map is presented as a LORETA image for CSD, typically representing the squared magnitude (or power, measured in $\mu A^2/(mm^4 \cdot Hz)$) of all computed current density vectors (with one vector of three components assigned to voxel) (Anderer et al., 2001). In addition, the linearity of LORETA allows to easily separate the different spectral characteristics of the generators, thus, providing CSD maps specific to each selected frequency band (Pascual-Marqui et al., 2011).

Given the existence of an infinite range of solutions to the inverse problem (i.e., an infinite set of different distributions for the current generators that can successfully reproduce the exact measured electrical signals), LORETA aims to constrain them until obtaining the single one with the property of maximal spatial smoothness (Pascual-Marqui et al., 2002). In technical terms (to be later simplified), a precise definition is that LORETA minimizes the squared norm of the Laplacian of the weighted three dimensional current density vector field (Del Felice et al., 2014). Thus, the LORETA solution is achieved by imposing a distribution for the current generators satisfying maximum similarity between nearby voxels (hence, aiming to reproduce a realistic neurophysiology of local clusters of synchronized cells), in terms of the module (strength) and spatial orientation of the current vectors representing neuronal activity (Lantz et al., 1997). Therefore, LORETA computes current vector fields for the electrical activity originated by neuronal populations associated to each voxel, which the property that the difference between currents corresponding

to a small sphere of voxels centered on a particular voxel has minimum difference with the voxel located in the center (which is a more precise mathematical meaning of “smoothness” for the LORETA solution) (Pascual-Marqui et al., 1994). However, by incorporating the smoothness principle in the spatial distribution of sources, LORETA produces some blurring in the calculated solution, which is more obvious for deep sources due to their increased difficulty for precise localization (Pascual-Marqui et al., 2002, Grech et al., 2008).

Alternatively, LORETA can also find a solution that, instead of calculating current vector fields, fixates from the beginning the orientation of electrical currents according to a template of the cortical surface of the brain, and with this anatomical constraint, calculates the solution that maximizes smoothness only for the modules of the vectors (i.e., spatial correlation between nearby voxels) (Seeck et al., 1996). LORETA solutions are also equally adept to be applied either in the time or frequency spaces (via Fourier transformation) (Pascual-Marqui, 2002).

In addition of computing a solution for the current generators explicitly aiming to be compatible with the known neurophysiological basis, LORETA is superior to the rest of inverse solution methods because it achieves true zero localization error (Pascual-Marqui, 2002, Jatoi et al., 2014). Indeed, since the purpose itself to create a source localization method is to be as precise as possible pinpointing brain structures that are active during a task, minimizing the localization error is arguably the most important criterion to adopt one particular neuroimaging method as compared to alternative approaches (Pascual-Marqui et al., 1994). Localization error and its minimization refers to the problem of producing an algorithm aiming to calculate a solution for the putative localization and activity levels of the generators fulfilling the condition of having the smallest possible difference when compared to the actual arrangement and activity levels of the real generators (Pascual-Marqui et al., 2002). It has been proven (using diverse computational models for comparison) that LORETA has the virtue of outcompeting all additional linear algorithms (e.g., weighted minimum norm method), and that its superiority becomes even clearer in the localization of deep

sources in the brain, generally doing so with virtually zero localization error (margin of error within only one voxel on average) (Lantz et al., 1997, Pascual-Marqui et al., 2002, Grech et al., 2008, Park et al., 2015).

Besides a robust computational and theoretical grounding, there exist a considerable number of studies that have validated the excellent accuracy of the LORETA method at the experimental level (Pascual-Marqui et al., 1994, Seeck et al., 1996, Jatoi et al., 2014). Empirical comparisons build upon the fact that the most critical brain sources linked to a given task are already known (in many cases, with an outstanding precision) from previous research (typically, by using fMRI), therefore, making them suitable for direct testing (i.e., comparing the relevant brain sources involved in an activity with the ones yielded by LORETA, completely on its own) (Lantz et al., 1997). These studies comprise the correct localization of: auditory stimuli (including identification of mismatch negativity for grammatical errors in spoken language and activity of Wernicke's area in semantic tasks, (Waberski et al., 2001)), visual and visuo-motor tasks (including facial recognition in the fusiform gyrus, (Pizzagalli et al., 2000)) and motor tasks (located in motor cortices, (Gomez et al., 2003)) during event-related potential (ERP) experiments (Pascual-Marqui et al., 2002, Jatoi et al., 2014). Furthermore, in patients with known epileptic foci, LORETA could correctly confirm the source of seizures that was diagnosed by trained neurologists (using either structural MRI or intracranial recordings) (Pascual-Marqui, 2002). Psychiatric diseases showing abnormal brain activity in previously recognized areas have also been used for testing, e.g., depression (exhibiting heightened activity in rostral anterior cingulate cortex), ADHD (pinpointing to the prefrontal cortex, which regulates attention and impulse control) and Alzheimer's disease (with LORETA generators indicating the same affected structures than PET scans) (Pascual-Marqui, 2002, Pascual-Marqui et al., 2011).

Finally, regarding biomedical applications derived from studies of source localization which underscore its importance, besides theoretical motivations, we can mention: diagnoses of mental diseases (e.g., schizophrenia, impulsivity, anxiety, etc.), localization of brain

tumors or foci of epileptic seizures, assessment of damage to brain tissue (e.g., after stroke or concussion) and other pathologies producing aberrant EEG patterns (e.g., excessive delta activity during wakefulness) (Esslen et al., 2008b, Gianotti et al., 2009, Ikeda et al., 2015).

## 2.3. Applications of LORETA to sleep

*"To understand sleep, we try to reformulate biological questions in terms of mathematics, typically systems of differential equations."* Janet Best[2.3]

Connemann et al. calculated CSD maps using LORETA for assessing intracranial current densities on a young healthy subject (with a previously known alpha-delta sleep pattern), covering from wakefulness and sleep onset (SO) to slow wave sleep (SWS) and REM sleep (first cycle) (Connemann et al., 2001). Recordings employed 19 electrodes distributed according to the international 10/20 system. LORETA maps were computed in the frequency domain and for the most important frequency bands. The delta band was defined as 1.5-6 Hz, the theta band as 6.5-8 Hz, the alpha band as 8.5-12 Hz, and the beta band as 12.5-30 Hz. CSD maps (representing spatial distributions of power spectral energy) were log-transformed for each band and statistically compared between different sleep stages. Statistics were based on calculation of *t*-tests (as a measure of statistical distance, similar to effect size), corrected for multiple comparisons (since there are thousands of voxels in LORETA, as in fMRI) using the statistical non-parametric randomization method (SnPM) (based on bootstrapping and surrogate randomization; 5,000 surrogates were generated in order to achieve optimal precision) (Nichols and Holmes, 2002). In this way, statistical maps can be created with LORETA (besides CSD maps), where each voxel has an associated *P*-value, with a threshold for significance at $P<0.05$.

During wakefulness, a preliminary result based on total spectral power (i.e., corresponding to the integral of energy for all frequency bands) revealed Brodmann area (BA) 18, in the occipital cortex, as the most active region, that is, where total power reached an absolute maximum

(Connemann et al., 2001). After discounting the absolute maximum, local maxima corresponding to BA 19 (a critical visual association area), BA 39 (angular gyrus) and BA 32 (dorsal anterior cingulate gyrus) also exhibited heightened activity during wakefulness. Then, a frequency-specific analysis revealed that theta, alpha and beta power were much higher during waking than during any sleep stage among occipital areas (in particular, alpha power being the more intense). Entering sleep, LORETA validated previous findings indicating progressive diminished alpha activity in the occipital lobe accompanying the SO transition (paralleling a reduction of metabolic consumption on the visual cortex at the neurophysiological level) (Wright et al., 1995). Following SO, during light sleep, alpha activity increased in the left frontal lobe, anterior and medial right frontal lobe, and anterior and parietal cingulum.

During stage 2 NREM sleep, total power underwent an increase in parietal areas compared to waking, although it decreased in occipital areas (there was no statistical difference in bilateral temporal cortices compared to wakefulness). In the frontal cortex, delta power increased intensely, but it was also paralleled by a concomitant reduction of beta power (hence, total power from 1.5 to 30 Hz did not vary significantly compared to earlier stages). The maximum statistical difference in total power during stage 2 NREM sleep compared to wakefulness was observed in BA 6 (paracentral lobule). Stage comparison of stage 2 with REM sleep revealed greater total power on the former. A frequency specific comparison revealed that, in the delta band, the whole brain showed greater activity than during either REM sleep or wakefulness (especially at the frontal pole). In addition, alpha and theta power were also higher during stage 2 throughout the whole frontal lobe, with peak differences located on BA 6, 8 and 9 (superior frontal gyrus and adjacent areas). Beta power was larger than in REM sleep in all cortical areas, except for the occipital cortex, possibly indicating higher beta activity during REM sleep associated with visual processing of dream content (Hobson, 2009, Nir and Tononi, 2010). When statistically compared to wakefulness, theta, alpha and beta power during REM sleep were significantly lower in occipital and temporal cortices, and significantly larger in the medial frontal cortex (maximum in BA 6 and 24). Nevertheless, activity in the parietal

cortex did not have a significant variation in any of these frequency bands in REM sleep compared to waking.

Later, during slow sleep (SWS), alpha activity was significantly larger than during wakefulness, displaying a topological distribution indicative of spatial heterogeneity in the process of falling asleep (Marzano et al., 2013). In addition, maximal CSD values in the alpha band were located symmetrically in the left and right cingulum (subcortical structures pertaining to the frontal lobe). Statistical comparison of SWS with REM sleep, stage 2 and wakefulness, revealed greater total power throughout all cortical areas in every comparison. Only exceptions include the medial parietal cortex, where total power was smaller than during stage 2, and the occipital cortex, displaying lower power than in waking. In the delta band, CSD was systematically larger than in any other frequency band throughout the entire cortex, except for a small region in the post-central gyrus (located in the lateral parietal lobe), showing no power difference with the previous stage 2. Frequency specific analyses revealed that alpha power was greater than during stage 2 in the left occipital, right middle frontal gyrus and parietal cortices. Beta power was generally lower than in stage 2, as SWS reflects a deeper sleep stage and greater disengagement from the external environment. Alpha and theta power were lower than in wakefulness for occipital areas, yet, frontal areas (extending to the temporal pole) showed a significant increase on these bands. In particular, highest alpha power during SWS was located on BA 32 and 24 (frontal and parietal limbic lobe).

Finally, in REM sleep, total power was lower than in wakefulness, except for the paracentral lobule and nearby regions (possibly because of heightened motor activity associated with dreams) (Hobson and McCarley, 2003, Zoetmulder et al., 2016, Siclari et al., 2017). Alpha, theta and beta power during REM sleep were also lower than in waking, with a maximum of activity in BA 6 (medial frontal gyrus), and secondly, BA 24 (cingulate gyrus). Frontal delta power was lower than in wakefulness (which is compatible with the subjective aspect of not noticing illogical or unrealistic scenarios characteristic of dreams), with no statistical difference in the occipital cortex (compatible with

visual processing of dreams) and higher power in the parietal cortex (as the parietal lobe is a major association hub, perhaps explaining the bizarre associations of dreams) (Hobson and McCarley, 2003).

In summary, during wakefulness, alpha power prevailed in occipital areas. With SO, alpha power decayed in intensity, but continued to be diffusively spread throughout the cortex. With increased sleep depth, alpha power increased in the frontal lobe and cingulum, and in SWS, frontal alpha was much larger than in wakefulness, with maximum CSD activity located on BA 32 and 24 (bilateral anterior cingulum). Delta power increased across all fronto-parietal areas and anterior limbic lobe as sleep progressed to SWS. Finally, CSD activity in the alpha-delta range tended to shift posteriorly following a fronto-occipital power gradient at the transition to REM sleep.

Anderer et al. used LORETA to localize the origin of sleep spindles (further subdivided between slow and fast spindles) and tracked their dynamical emergence with the SO transition (Anderer et al., 2001). The solution space was restricted to cortical gray matter and hippocampus (given the relationship of this subcortical structure with spindles) (Andrillon et al., 2011, Sarasso et al., 2014). In particular, subcortical structures anatomically adjacent to the dorsal thalamus (where sleep spindles are generated) were investigated in closer detail, as LORETA proved to be successful at detecting the associated CSD signals (despite the relatively deep origin) (Magnin et al., 2010, Del Felice et al., 2014). It was hypothesized that slow and fast spindles would have different source origins, revealing at least two functionally separated spindle generators. Ten healthy young adults were recorded during a whole night. EEG acquisition employed 18 electrodes, spatially arranged according to the 10/20 system. Epochs of 1.25 s containing either frontal or parietal spindles were marked by visual inspection. A control condition of artifact-free epochs of 1-3 s (either anteceding or following a particular spindle event) was also introduced for statistical comparison.

The interest for introducing an inverse solution method in the study of spindles comes from the fact that scalp analysis might yield contradictory results when changing the EEG reference. For example, a mastoid versus reference recordings, slow spindles appeared more

distributed throughout anterior regions, and fast spindles, posteriorly (Zeitlhofer et al., 1997). In contrast, bipolar recordings showed predominance of slow spindles in posterior parts of the scalp, with fast spindles located on anterior derivations (Werth et al., 1997). Given that reference dependence produces results that are inconsistent and ambiguous, scalp power maps are useless in terms of inferring the actual underlying brain electrical generators, thus, imposing the necessity of a tool such as LORETA.

A topographical analysis via frequency decomposition (defined in the range of 7-14 Hz) during stage 2 NREM yielded a pattern of spindle recurrence every 3-10 s on the cortical surface, with each spindle lasting 1-3 s. All LORETA images were calculated as the average CSD map across subjects, corresponding to the matching epochs (either containing spindles or not). Thus, application of LORETA in the frequency domain on the spindle range, covering each consecutive bin from 11.2 to 15.2 Hz (frequency resolution of 0.8 Hz), confirmed the distinction between slow and fast spindles, as characterized by maximum power spectral density peaking at around 12 and 14 Hz, respectively. In order to define local maxima in a three-dimensional space of CSD voxel optimization, the algorithm searched for candidate voxels in which deviations along the three spatial dimensions CSD values did barely change, and in that case, would lower the values of the candidate voxels.

In the next step, a statistical comparison of LORETA maps between epochs containing spindles and control epochs (i.e., showing no spindles) was conducted. As part of LORETA image preprocessing, CSD maps were normalized and log-transformed, given that the power scale in the spindle condition covered a very broad range. Then, paired *t*-tests were performed in every frequency bin between 0.8 and 24 Hz (i.e., from slow delta up to the beta band). The statistical threshold at $P<0.05$ was based on the method of statistical non-parametric randomization (SnPM, generating a total of 5,000 surrogates), which automatically corrects for multiple comparisons, with the null-hypothesis of no difference between spindle and control epochs rejected even if only one voxel crossed the threshold (Nichols and

Holmes, 2002). Consequently, voxels whose *t*-values did not reach the critical statistic threshold corresponding to $P$=0.05 could be eliminated, with the remaining voxels considered as defining clusters of regions of significant spindle activity.

As a result of these statistical tests, spindle activity emerged as significant in the bilateral and medial frontal lobe, and medial parietal lobe (hence, confirming the existence of multiple cortical spindle sources). Furthermore, hemispheric asymmetries could be detected, showing a small, yet higher, spindle prevalence on lateral left hemisphere (this asymmetry was specifically attributed to structures related to spindle generation, not to inherent functional hemispheric differences on NREM sleep itself). In addition, when parsing between fast and slow spindles, it was observed that slow spindles (upper limit of 13 Hz) had their maximal CSD values in BA 9 and 10 (frontal pole and nearby regions), whereas fast spindles (lower limit of 13 Hz) displayed their maximum in BA 7 (precuneus, a structure in front of the occipital cortex). A proposed neurophysiological explanation for the finding of these particular brain areas (known to contribute to the processing of sensory stimuli) involves their anatomical connections with thalamic nuclei (where spindles are firstly generated), thus, allowing an axonal path for propagation. In this way, BA 9 and 10 have principal white matter projections to the dorsomedial thalamic nucleus, and BA 7 connects to the latero-posterior, latero-dorsal and rostral centro-lateral thalamic nuclei (Magnin et al., 2010). The localized brain sources were found to be simultaneously active during most spindle episodes (with only minor exceptions in which only either frontal or parietal cortices were active).

Brodmann areas indicated by LORETA as significant during spindle episodes have in common a known importance in the neuronal processing of sensory data (Andrews-Hanna et al., 2010). The precuneus is a major hub for the integration of information coming from diverse cortical sensory areas, hence, allowing for further processing of sensory input (Vogt and Laureys, 2005). On the other hand, the prefrontal cortex and anterior cingulate cortex are relevant for higher cognitive functions and decision making (Rogers et al., 1999). Therefore, a possible neurophysiological interpretation is that,

compatible with what is known about spindles, the appearing brain areas exert a gating function that prevent sensory signals coming from the thalamus being relayed to the cortex, hence, assuring the continuity of sleep (Ogilvie, 2001). In addition, frequency differences between slow and fast spindles were attributed to differences traceable to the rate of neuronal oscillations between thalamic neurons projecting to the precuneus compared to those projecting to medial prefrontal cortices (the former being faster than the latter) (Kim et al., 1995, Anderer et al., 2001).

More recently, Del Felice et al. also investigated spindles using both an updated and improved version of LORETA and high-density EEG recordings (Del Felice et al., 2014). High-density EEG recordings with 256 channels were analyzed. Hence, different from traditional EEG caps (as in the previous study by Anderer et al.), a higher spatial resolution allows for a more reliable reconstruction of the generators by application of LORETA. Recordings were obtained in 18 healthy young subjects during daytime napping. Spindles were visually scored, and band-pass filtered between 10-12 Hz (for slow spindles) and 12-14 Hz (for fast spindles) for independent analysis. Artifact free segments containing either slow or fast spindles were separately averaged, both among subjects and corresponding frequency bins.

Thus, LORETA solutions identified, in a consistent fashion across segments and subjects, two to three slow spindle generators in frontal lobes. Additional sources included (in half of the cases) parietal and limbic lobes. In four subjects, generators for slow spindles appeared exclusively on the frontal lobe, although the sole activation of this lobe was not normative among participants. The most salient frontal areas (sorted in order of importance) included: BA 5, 6 and 11 (together constituting the medial prefrontal cortex and adjacent regions) and BA 45 and 46 (inferior frontal gyrus and dorsolateral prefrontal cortex). Thus, the results are in agreement with results from Anderer et al. in the importance of the prefrontal cortex and nearby areas, (Anderer et al., 2001). However, besides the frontal lobe, other statistically significant areas associated with slow spindles originated from this study: BA 31 (precuneus, part of the superior parietal lobule), BA 40

(inferior parietal gyrus) and BA 20 (uncus, pertaining to the parahippocampal gyrus) ($P<0.05$ based on $\chi^2$-tests in all cases). Moreover, average CSD in the slow spindle range from these posterior structures was not statistically different from main areas found in the frontal lobe.

On the other hand, LORETA solutions showed for fast spindles the presence of multiple temporal and parietal sources, with secondary sources of activity detected in frontal regions (however, for six subjects, there were no significant contributions from frontal structures). Thus, several spindle sources previously found for slow spindles in the frontal lobe also remained significant in the fast spindle range, including BA 6, 10 and 11 (medial prefrontal cortex and inferior frontal gyrus). Nevertheless, fast spindle activity was additionally coupled with activation of parietal and temporal structures (in agreement, but also widening, the work of Anderer et al. (Anderer et al., 2001)). Hence, significant areas (sorted in order of importance) included: BA 20 (inferior temporal gyrus), BA 21 (middle temporal gyrus), BA 40 (inferior parietal lobule) and BA 43 (parietal operculum). In three of the participants, an additional cortical generator of fast spindles located in a different lobe, BA 18 (lingual gyrus, in occipital cortex) emerged. Furthermore, a statistical comparison revealed that slow spindle sources, compared to the fast ones, were preferentially located on frontal structures ($P<0.001$). Both inter-individual and intra-individual differences in source intensities of slow and fast spindles were not statistically significant, implying that intensity of slow and fast spindles were comparable and similar among subjects.

Hence, slow spindles appeared preferentially over anterior structures, whereas fast spindles tended to be more scattered throughout the cortex, but showing maximal predominance in central structures. The pivotal role played by frontal areas (mainly BA 10 and 11) in spindle activity is explained, in neuroanatomical terms, due to the rich mutual projections existing between frontal areas and the thalamus (Rey et al., 2007). Overlapping areas displaying common generators for both slow and fast spindles were found in the medial and inferior frontal gyri, whereas the remaining generators were independent. Overall,

LORETA findings of multiple restricted cortical generators of spindle activity is in agreement with a previous study performed by Nir et al. (Nir et al., 2011). Yet, the aforementioned research was investigated in an invasive manner, by means of intracranial recordings obtained from epileptic patients during SO, hence, with the potential caveat of casting doubt on the generalization of its conclusion to a neurologically healthy group. Therefore, even when the LORETA solution was based on far field potentials recorded with the scalp EEG, results obtained from this study offer the same localization as invasive near field potentials recorded with stereo-EEG, thus, indicating not only the reliability of LORETA solutions, but also its much more practical use, as not relying on any invasive recording method.

Saletin et al. combined brain source localization provided by LORETA with structural MRI (a tool sensitive to gray mater differences among participants) to investigate structural brain correlates of sleep oscillations, and furthermore, the extent to which inter-individual variability in sleep features can be attributed to neuroanatomical differences based on certain key morphological characteristics (Saletin et al., 2013). The rationale for this study comes from the empirical observation that EEG features (e.g., spectral profile) within an individual are very stable from one night to the next, whereas there are often striking differences appearing between individuals in the same EEG features (partially attributable to genetically determined factors) (Finelli et al., 2001a). The most prominent differences of these EEG features (analogous to neuronal signatures) relate to sleep spindles and slow waves; hence, these bands were selected for correlational studies with morphological features indicative of singular brain structures (as can be revealed by LORETA maps), another main characteristic expected to be individually stable, but inter-individually different (Park and Friston, 2013). To this aim, relevant brain regions can be surmised based on the brain functions they support in relation to sleep, as well as with the detection of statistically significant source generators appearing during NREM sleep (indicated by LORETA).

EEG recordings were based on 19 electrodes (mastoid reference), arranged according to the international 10/20 system, with a total of 22

participants. In order to have a consistent temporal frame for statistical analysis, LORETA images were calculated around the instant of the negative peak of each slow wave (filtered between 0.5 and 4 Hz and detected at the Fz electrode). Thus, event-related potential (ERP) containing slow waves comprised 1-s epochs, centered at the negative peak (i.e., with 500 ms on either side). Each epoch was then averaged, firstly, within-subjects, and secondly, between-subjects, hence, producing a single grand average for the group, subsequently analyzed by LORETA for source localization in the time domain. Artifact free epochs containing spindles (averaged from midline electrodes Fz, Cz and Pz) were band-pass filtered between 11-13 Hz for slow spindles, and 13-15 Hz for fast spindles.

One key finding of this study was that gray matter volume in interoceptive and exteroceptive cortices (represented by the auditory cortex and somatosensory cortex, respectively) was correlated with lower density of slow frequency spindles during NREM sleep. In particular, gray matter volume in bilateral auditory cortices (including Heschl's gyrus), as well as anterior and posterior insula, correlated with slower spindle frequencies. Moreover, with a more lenient statistical threshold, additional temporal and frontal regions were also linked to slower spindle frequencies. Overall, these correlations support the role played by spindles in sleep efficiency by preventing conscious arousal, and consequently, assure sleep continuity (De Gennaro et al., 2000a, De Gennaro and Ferrara, 2003).

A second major finding was that the amount of gray matter volume in bilateral hippocampus (specifically, the more posterior regions) was positively associated with faster spindle frequencies during NREM sleep. The proposed biological interpretation was that greater gray matter volume in the hippocampus is consistent with its well-known role for declarative memory processing, with faster spindles allowing more efficient transferring of learned information to the cortex for long-term storage (Rasch and Born, 2013). Indeed, inter-individual differences in fast frequency spindles predict hippocampal dependent learning ability and fast frequency spindle activation temporally coincides with hippocampal activation (Lavenex and Amaral, 2000). In addition, gray matter volume in the parahippocampal gyrus also

predicted (through linear regression) inter-individual differences in slow wave density.

Thirdly, gray matter differences in basal forebrain and hypothalamus (two main brain regions involved in homeostatic sleep regulation, (Kalinchuk et al., 2010)), together with medial prefrontal cortex and insula, explained individual differences in slow wave density of NREM as they were positively correlated. The basal forebrain controls changes in slow wave density following prolonged wakefulness, whereas medial prefrontal cortex and insula account for individual differences in power of slow waves (Sarter et al., 2003). LORETA results were compatible with previous studies of source localization that have pinpointed to the cingulate cortex, an area greatly communicating to and adjacent to the medial prefrontal cortex, hence, consistent with its role as a generator of slow waves, both functionally (by EEG source localization) and structurally (by gray matter volume) (Murphy et al., 2011b, Sämann et al., 2011). Furthermore, gray matter volume in the basal forebrain accounted for approximately 50% of the inter-individual variability in slow wave density.

Finally, gray matter volume in the orbitofrontal cortex (a prefrontal region) and middle cingulate cortex were predictive of inter-individual differences in slow wave activity (SWA), with higher SWA related to larger gray matter volume in these structures. In this way, the prefrontal cortex is not only the region producing the highest SWA, but also the most relevant cortical area to explain inter-individual variability observed in the NREM sleep EEG (including homeostatic rebound after sleep deprivation), thus, acting as a kind of neuronal “fingerprint” (De Gennaro et al., 2005a, Dang-Vu et al., 2008). Additionally, discrete frontal and temporal regions were found to positively correlate with SWA, whereas occipital areas had a negative correlation. Besides the cerebral cortex, gray matter volume in the cerebellar cortex was positively correlated with slow wave density, but negatively with SWA.

Park et al. conducted an EEG study to investigate the SO transition with LORETA (Park et al., 2015). To this aim, 61 young healthy participants were recruited and recorded with 32 electrodes (following

the 10/20 international system, mastoid reference). The authors did not define SO as the moment where the first spindle or K complex was detectable, but as the entering into stage 1 NREM sleep. Thus, four types of 30-s epochs were analyzed: wakefulness, the transition stage, early stage 1 and late stage 1. At the EEG level, the waking state was determined by epochs exhibiting high alpha power in occipital channels O1 and O2. Then, the transition stage comprised the last 15 s of wakefulness and the first 15 s of stage 1. Early stage 1 started at the end the transition stage and was characterized by strong theta power. Finally, late stage 1 was defined from the end of early stage 1 until the final stage 1 epoch.

EEG data corresponding to these epochs was transformed to the frequency domain to calculate the associated cross-spectral matrices. LORETA images were computed upon previous cross-spectral results, hence, producing several CSD maps representing 3-D power distribution in a set of relevant frequency bands, yielding one CSD map per band. The frequency bands included for analysis were: delta (1-3.5 Hz), theta (4-7.5 Hz), alpha-1 (8-10 Hz), alpha-2 (10-12 Hz), beta-1 (12.5-18 Hz), beta-2 (18.5-21 Hz) and beta-3 (21.5-30 Hz). Statistical comparisons between the four different defined states rejected the null hypothesis of no difference if more than 50% of voxels in a particular cortical gyrus or Brodmann area displayed a significant change. In this way, three tests were performed, representing consecutive contrasts: transition stage versus wakefulness, early stage 1 versus transition stage and late stage 1 versus early stage 1. The method of choice was statistical non-parametric randomization (or SnPM), and the statistic employed was the logarithmic ratio of CSD averages (based on Fisher's *F*-value) for comparison between conditions. A total of 5,000 surrogates were generated for all contrasts, with a level of statistical significance of $P<0.01$ (corrected for multiple comparisons by SnPM).

In the first test (transition stage versus waking), LORETA results showed increased delta and theta activity in the occipital lobe (maximal increase in BA 18), accompanied by a significant reduction in alpha-1 activity (maximal decrease in BA 19). Theta activity increased in the bilateral parietal and occipital lobes, posterior region

of the frontal lobe (including the precentral gyrus and paracentral lobules) and parts of the limbic lobe (anterior and posterior cingulate cortices and parahippocampal gyrus). Highest theta increase entering into the transitional stage occurred in BA 23 (ventral posterior cingulate gyrus).

Later, in the second test (early stage 1 versus transition stage), there was a global decrease in alpha power (comprising both alpha-1 and alpha-2), with exceptions found in anterior frontal regions (superior and middle frontal gyri and orbitofrontal cortex for alpha-1 and orbital gyrus for alpha-2). The greatest decrease in alpha activity (both for alpha-1 and alpha-2) were located in BA 27 (parahippocampal gyrus) and BA 23 (cingulate gyrus). Theta activity increased throughout most occipital and parietal areas and, additionally, in the limbic lobe (maximal change in the posterior cingulate cortex) and posterior frontal lobe (maximal change in the paracentral lobules). The highest theta increase happened in BA 31 (posterior cingulate gyrus).

Finally, in the third test (late stage 1 versus early stage 1), beta-2 and beta-3 decreased in the whole frontal lobe and partially in the parietal and temporal lobes. In particular, beta-2 decreased in frontal (superior and middle frontal gyri and precentral gyrus) and limbic areas (parahippocampal gyrus and the whole cingulate gyrus), including other structures such as the insula, transverse temporal gyri and inferior parietal lobules. Beta-3 decreased along a greater cortical extension than beta-2: frontal lobe (superior and middle frontal gyri, precentral gyrus and medial frontal gyrus), limbic lobe (uncus, whole cingulate gyrus and parahippocampal gyrus), parietal lobe (postcentral gyrus, inferior parietal lobule, superior parietal lobule and precuneus) and the whole temporal lobe. Beta-2 and beta-3 had their greatest reduction in BA 24 (cingulate gyrus). On the other hand, alpha-2 decreased in the parietal lobe (postcentral gyri and inferior parietal lobules), frontal lobes (paracentral lobules) and limbic areas adjacent to the occipital lobe (cingulate gyri, posterior cingulate cortex and precuneus). Greatest alpha-2 decrease occurred in BA 23 (cingulate gyrus).

In summary, a three step process is proposed, starting with an increase in theta waves in posterior regions of the brain, followed by a global

reduction in alpha power and concluding in a beta decrease in fronto-central areas. Hence, there is an ordered pattern in the gradual progression of frequency bands exhibiting significant changes in CSD activity throughout the cortex, starting from slower frequency bands (theta) and ending with a preponderance of statistical changes in faster frequency bands (alpha and beta). In addition, spectral changes began in posterior areas, and then recruited most of the cortex, with increasing fontal and parietal predominance. In the alpha band, initial prevailing posterior reduction, culminating in a global decrease, is consistent with the notion of alpha power representing inhibition of task-irrelevant areas during tasks requiring executive function control (called inhibition-timing hypothesis, (Klimesch et al., 2007)) as SO implies that any potential ongoing cognitive process demanding engagement with the environment will gradually fade (Wright et al., 1995, Ogilvie, 2001). Finally, the overall decrease of beta activity during the SO transition is supposed to reflect the progressive decrease of muscle and sensory-motor activity (as shown by the fading of beta activity in central areas), as well as alertness levels (evidenced by the frontal vanishing of beta activity), suggesting isolation from the external environment and, thus, becoming more dependent on internal dynamics.

Bersagliere et al. performed a detailed study for delta activity through power spectral maps and source localization based on exact LORETA (eLORETA) (Bersagliere et al., 2018). The delta band was split between a low delta (0.5-1 Hz) and mid-delta (1.25-2 Hz) component for separate analyses. This split was performed as the results were qualitatively different based on this distinction. Foremost, only the mid-delta component was sensitive to effects of sleep deprivation (by increases in power), while the low delta component was not affected. Power maps during the first NREM sleep episode revealed increased activity in frontal regions in both components (additionally, there was an occipital focus of activity in the low delta range). LORETA revealed a frontal distribution of activity in both delta components, with sleep deprivation affecting fronto-parietal sources by significantly increased activity for the mid-delta component (with 19.5% of all voxels significantly increasing). Irrespective of the level of sleep pressure, occipital and temporal areas had higher activity in the low

delta component (there was an asymmetry towards the right hemisphere), whereas in the mid-delta component, limbic and frontal sources were more predominant (with no discernible asymmetry). Additionally, sleep deprivation was found to influence inter-hemispheric asymmetry, with higher spreading in left frontal areas (39.2% on the left compared to 13.2% on the right frontal cortices, expressed normalized to the total number of significant voxels).

## 3. Brain dynamics during the sleep onset transition: an EEG source localization study


### Abstract

EEG source localization is essential to understand, e.g., which brain areas are implicated in cognitive tasks and basic brain functioning. We applied LORETA, a technique of EEG source localization, to identify the principal brain areas involved in the process of falling asleep (sleep onset, SO). We localized the contributing areas and tracked their temporal evolution (in 2-min intervals from 2 min prior to SO up to 10 min after SO). We focused on delta (0.5-5 Hz) and sigma activity (12-16 Hz; spindles) during a baseline night and subsequent recovery sleep after total sleep deprivation of 40 hours.

Delta activity gradually increased both in baseline and recovery sleep, starting in frontal areas and finally involving the entire cortex. This increase was steeper in the recovery condition. The evolution of sigma activity resembled an inverted U-shape in both conditions and the activity was most salient in the parietal cortex. In recovery, sigma activity reached its maximum faster than in baseline, but reached lower levels. The dynamics in recovery could be considered as a “fast-forward version” of the one in baseline in both frequency bands due to higher sleep pressure (homeostasis).


**Key words:** EEG sources, transition into sleep, homeostasis, delta activity, sigma activity, LORETA.

## 3.1. Introduction

*"On the walls of the cave, only the shadows are the truth."* Plato[3.1]

The process of falling asleep (sleep onset, SO) may be considered as multi-dimensional, entailing, e.g., subjective, behavioral and physiological dimensions (see Fig. 3 in (Ogilvie, 2001)). Frequency specific systematic EEG changes during wake-sleep transitions have been reported (Hori, 1985, Merica and Gaillard, 1992, Ogilvie, 2001). In general, a pattern of declining alpha and theta activity accompanied by increased delta activity has been observed. Not all brain areas exhibit the same frequency-specific EEG changes and not all regions change simultaneously during the wake-sleep transition (Hori, 1985, Wright et al., 1995, Berry, 1996, Ogilvie, 2001, Ferrara and De Gennaro, 2011). While occipital alpha activity is predominant in waking, the transition to sleep is accompanied by a posterior reduction of alpha activity; nonetheless, following SO, alpha activity spreads towards more anterior areas (De Gennaro et al., 2004). Overall, the progressive synchronization of the EEG during the SO transition is reflected, firstly, in a fronto-parietal prominence of low frequency activity, secondly, in a maximum of spindle activity in centro-parietal areas, and finally, in a shift of alpha activity towards more anterior regions (Ferrara and De Gennaro, 2011).

An important caveat in studying the transition to sleep from waking is the particular criterion or definition adopted to operationalize SO based on EEG recordings in humans. Many definitions have been proposed and applied in the sleep literature; nevertheless, no general consensus has yet been established (Wright et al., 1995, Ogilvie, 2001, De Gennaro et al., 2004, Picchioni et al., 2008, Parrino et al., 2009, Prerau et al., 2014, Speth and Speth, 2016). The American Academy of Sleep Medicine (AASM) defines the moment of SO as the first appearance of any 30-s epoch containing at least 15 s of any sleep stage sleep (Iber et al., 2007a). On the other hand, the Rechtschaffen and Kales manual defined sleep onset by the first three consecutive epochs of stage 1 (or any other deeper stage of sleep) (Rechtschaffen and Kales, 1968).

Prerau et al. (Prerau et al., 2014) considered SO to be a continuous dynamic physiological process. They proposed a series of biological markers that can be mathematically modeled to track the SO process. These markers were the reduction of alpha power, the increase of theta and delta power and the reduction of muscle activity in a behavioral task (sustained breathing paradigm) while subjects were falling asleep. Combining these measures revealed that SO is gradual process reflecting decreasing levels of vigilance. Siclari et al. (Siclari et al., 2014) defined a time window for the SO process characterized by two different events delineating the beginning and end of the window. The beginning was associated with the disappearance of alpha activity, when it was replaced by the high frequency, low voltage activity typical of stage 1 of non-REM sleep. The end of the SO window was defined by the first slow-wave sequence (FSS, generally appearing at the transition between stage 2 and stage 3 of non-REM sleep). The first slow-wave burst in the falling asleep period had to consist of more than two successive slow waves (half-wave duration >0.5 s and >75 μV peak-to-peak amplitude of each slow wave) not followed by an arousal. They investigated SO through multiple awakenings across the night and the average length of a SO window defined in this way was 9.3±2.1 (SD) min.

Here, we considered the first occurrence of stage 2 (first emergence of a spindle or K-complex) as SO, as indicating unequivocal signs of sleep as also used in other studies (De Gennaro et al., 2005b, Magosso et al., 2007, Marzano et al., 2013, Sarasso et al., 2014, Vecchio et al., 2017). We assessed what happens during the wake-sleep transition at a cortical level, assessed by EEG recordings from the scalp. We applied EEG source localization (exact low resolution electromagnetic tomography, eLORETA, (Pascual-Marqui et al., 2002, Pascual-Marqui et al., 2011, Pascual-Marqui et al., 2014b) to identify the brain areas involved in the process of falling asleep. Using such a technique, it is possible to reconstruct grey matter sources that generate the observed EEG activity with maximal smoothness. The data were divided into intervals of 2 min, from 2 min prior to SO up to 10 min after SO, yielding six 2-min intervals (one before SO and five after SO). We

choose a 2-min interval length as this was the interval common to all participants prior to sleep onset in the recovery condition. Average LORETA images of different frequency bands were computed for these 2-min intervals to track the temporal evolution of the SO process. Furthermore, the impact of increased sleep pressure on the SO process was investigated.

## 3.2. Methods

Source localization was performed on an existing dataset of eight healthy young males (Finelli et al., 2000, Finelli et al., 2001b). The original study was designed to investigate the effects of sleep deprivation on EEG topography. Polysomnography was performed during an adaptation night, a subsequent baseline night, and a recovery night after 40 hours of sustained wakefulness. Twenty-seven scalp EEG electrodes (extended 10-20 system; (Finelli et al., 2001b, Rusterholz et al., 2010)) were recorded during baseline and recovery sleep. The EEG signals were sampled at 128 Hz (band-pass filter: 0.16 - 30 Hz; for additional details see (Finelli et al., 2001b)). Sleep stages were visually scored for 20-s epochs according to the standard criteria (Rechtschaffen and Kales, 1968). Artifacts were identified as described in (Finelli et al., 2001b). Additionally, epochs exhibiting discontinuities in the EEG signal or extreme departure from previous range of values (e.g., as caused by sweating or head movements) were also excluded. In case of 4-s epochs containing artifacts that could not be removed, they were replaced by the preceding epoch free of artifacts, as source localization was finally performed on 12-s epochs (i.e., three consecutive 4-s epochs). Originally, EEG data were recorded against a technical reference (reference electrode placed 5% rostral to Cz). For the subsequent processing they were re-referenced to the average reference.

The strength and distribution of the intracranial sources (squared magnitude of current source density (smCSD)) of scalp electric potential differences were estimated with eLORETA (exact low resolution electromagnetic tomography, (Pascual-Marqui et al., 2011, Pascual-Marqui et al., 2014a))) (see also Supplementary Methods, section 3.7.1). We applied implementation 20160611 (most updated at

the time of running this analyses) of eLORETA (available as free academic software from http://uzh.ch/keyinst/loreta). The solution space was restricted to cortical gray matter, corresponding to 6239 voxels with a spatial resolution of 5 mm. In order to compute the smCSD maps, the first step consisted of calculating cross-spectral matrices from the EEG (12-s windows). Then, the pseudoinverse rendered by LORETA transforms the frequency information at the level of the scalp to the underlying currents inside the brain at each voxel. As the analysis was performed on 2-min intervals, ten 12-s maps were averaged, finally rendering the CSD maps. The following frequency bands were analyzed: delta (0.5-5 Hz), theta (5-8 Hz), alpha (8-12 Hz), sigma (12-16 Hz) and beta (16-24 Hz). Although the paper focuses mainly on delta and sigma activity, the results regarding alpha and beta activity are reported in the Supplementary Material.

Statistical comparisons (2-min intervals, conditions) of LORETA images were based on statistical non-parametric mapping (SnPM, (Nichols and Holmes, 2002)). This approach is based on the non-parametric randomization of the maximum-statistic over all voxels, providing an estimator for the empirical probability distribution under the null hypothesis (“no difference between conditions”), which does not need to rely on the assumption of a specific or exact statistical distribution (e.g., Gaussian), and at the same time also accounting for multiple comparisons. To this aim, a total of 5,000 surrogates were generated in the application of the SnPM method. The temporal evolution of cortical sources was investigated by comparing consecutive 2-min intervals and the 2- min intervals after SO with the one prior to SO. First, on a global level we determined how many voxels have differed between 2-min intervals and between conditions (Figs. 3.1, 3.4, S3.1, S3.4, and S3.7). Second, smCSD maps of selected 2-min intervals were illustrated (Figs. 3.2, 3.5, S3.2, S3.5, and S3.8) and 2-min intervals were compared to assess the temporal changes of cortical sources underlying oscillatory activity, separately for the two conditions (Figs. 3.3, 3.6, 3.7, S3.3, S3.6, and S3.9). Differences between conditions were assessed at the global level, i.e. how may voxels have differed between the conditions and we report the 3 brain

regions that showed higher activations during recovery (Tables 3.1 and 3.2).

## 3.3. Results

We investigated the temporal evolution of smCSD (mean across voxels, maps) of different frequency bands (delta, theta, alpha, sigma -spindles-, and beta), i.e. the cortical sources underlying their activity, at the transition into sleep under baseline conditions and after 40 h of sustained wakefulness (recovery sleep). Two minutes prior to SO (first occurrence of stage 2) and five 2-min intervals after SO were investigated (see Methods).

In the baseline condition, participants fell asleep within 12.0±1.6 min (SEM) and significantly faster within 4.0±1.2 min in recovery sleep ($P<0.01$; Wilcoxon signed-rank test).

### 3.3.1. Temporal evolution of delta activity at sleep onset

Mean delta activity averaged across the brain (all 6239 cortical gray matter voxels) gradually increased after SO in both conditions (Fig. 3.1, left, blue and red line), with a steeper increase in the recovery condition. Prior to SO there was no difference in activation, however, after SO the percentage of voxels differing between the two conditions steadily increased reaching 100% at 9 min (8-10 min; Fig. 3.1, left, green line). The largest leap in statistical divergence between the conditions occurred at 3 min (2 to 4 min) after SO, >70% of the voxels showing higher activity in recovery sleep. Delta activity in both conditions showed a monotonic increase in the time window investigated (Fig. 3.1 left). Therefore, an exponential function ($\sim\exp(\alpha \cdot t)$; t, time) was fitted (Matlab function “fit” with ”exp1” revealing the 95% CI). The coefficient controlling the steepness of the exponential increase was $\alpha$=0.094±0.008 $min^{-1}$ for baseline and $\alpha$=0.144±0.022 $min^{-1}$ for recovery, revealing a significant difference as the 95% CI did not overlap.

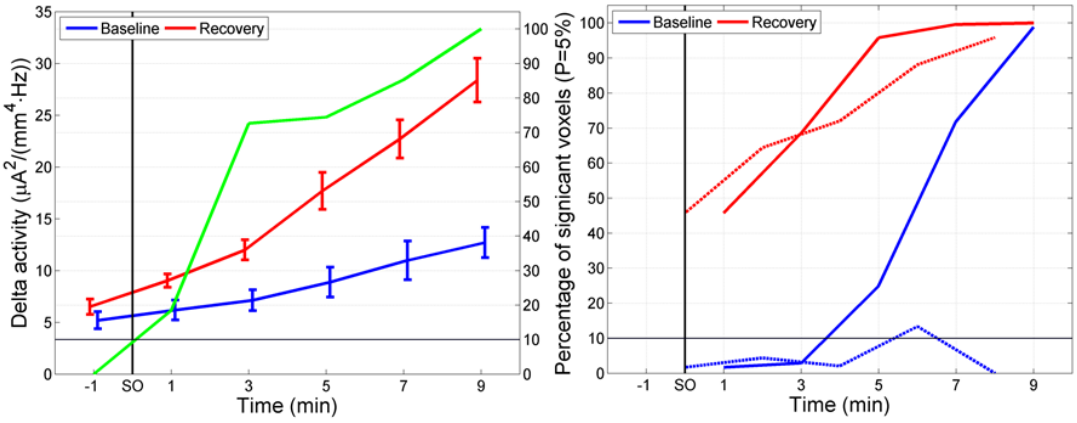


***Fig. 3.1.*** *Left panel: Temporal evolution of delta activity (mean squared magnitude of current source density in the delta band 0.5-5 Hz; [μA²/(mm⁴·Hz)]; averaged across all voxels; left y-axis) and statistical evaluation (percentage of significantly different voxels between conditions; right y-axis) at the transition to sleep (sleep onset, SO). Blue line: mean delta activity in baseline; red line: mean delta activity in recovery; green line: percentage of voxels significantly different between recovery and baseline. Right panel: Percentage of voxels significantly different from the 2-min interval prior to SO (solid lines) and percentage of voxels significantly different between consecutive 2-min intervals (dashed lines). Blue lines: baseline; red lines: recovery.*

The percentage of voxels differing from the 2 min prior to SO was considerably larger in the recovery condition than in baseline. Approximately 45% of the voxels differed already in the first 2-min interval after SO (Fig. 3.1, right, solid lines) whereas in baseline differences started to emerge 5 min after SO (4-6 min), indicating that delta activity increased faster in recovery than baseline. In the last 2-min interval, in both conditions, activity in all voxels was larger than before SO. Comparing consecutive 2-min intervals, there was basically no difference in baseline, however, during recovery large differences started to emerge immediately after SO (Fig. 3.1, right, dashed lines) again, indicative of a faster increase.

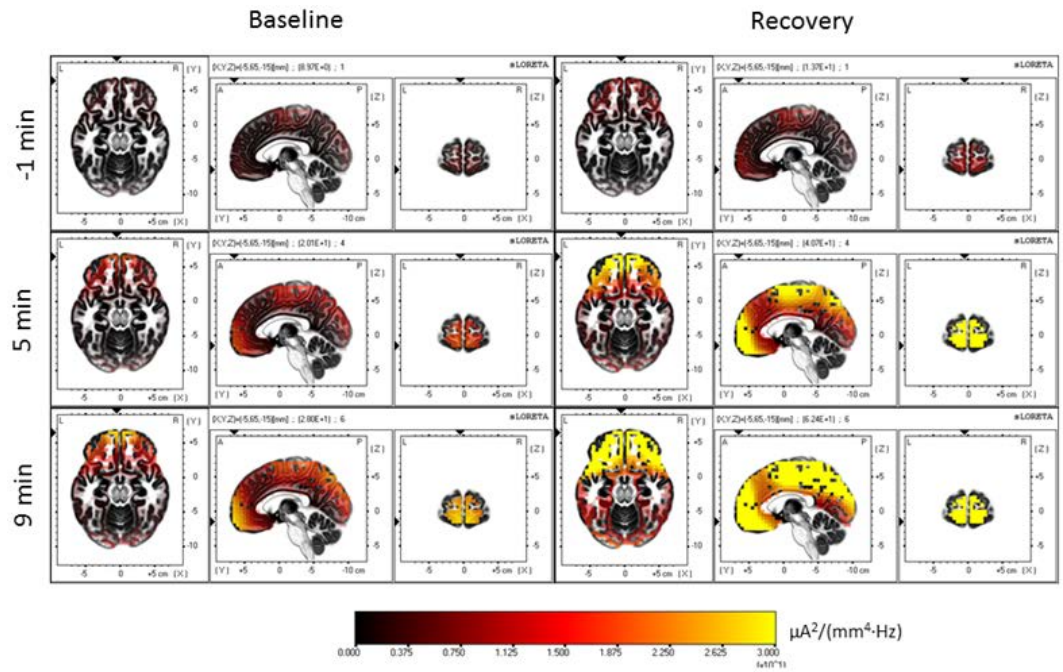


***Fig. 3.2.*** *LORETA source estimation of delta activity (squared magnitude of current source density [μA²/(mm⁴·Hz)]) in the delta band (0.5-5 Hz)). Left: baseline, right: recovery sleep. Images represent average values of the two minutes before SO (-1 min), 4-6 min (5 min) and 8-10 min (9 min) after SO. The plane coordinates were kept constant (-5, 65, -15; MNI coordinates [mm]), matching the areas showing the highest degree of delta activity during the SO transition.*

Fig. 3.2 illustrates the gradual spatial buildup of the cortical sources of delta activity (two minutes before the transition (-1 min), 4-6 (5 min) and 8-10 min (9 min) after SO), both, in baseline (left) and recovery sleep (right). To facilitate the temporal tracking, the plane coordinates at the different time points were kept constant matching the areas showing the highest degree of delta activity during the SO transition. Areas with highest delta activity were located in the frontal and the parietal lobe, specifically, in BA 11 (most of the prefrontal cortex; Table 3.1) and BA 3 (the postcentral gyrus). Delta activity in recovery sleep rose faster, including the parietal cortex, but the highest activity was still located in the frontal lobe, where the effects of sleep deprivation were most noticeable.

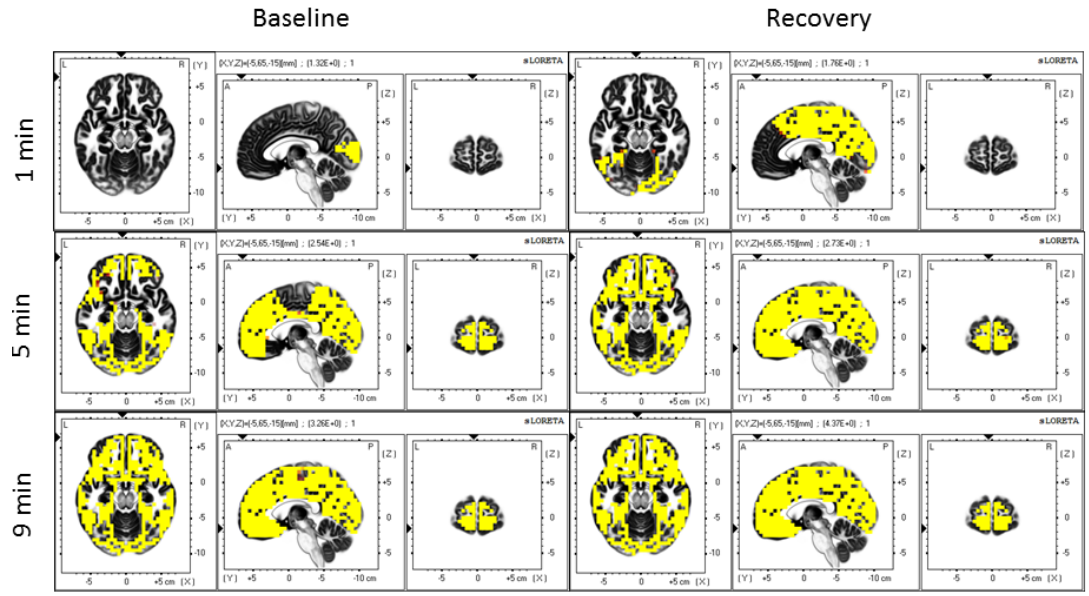


***Fig. 3.3.*** *Statistical evaluation of the temporal evolution of delta activity (0.5-5 Hz). Left: baseline, right: recovery sleep. Yellow areas represent voxels that showed higher delta activity after SO (t-values at P<0.05). Minutes 0-2 (1 min), 4-6 (5 min) and 8-10 (9 min) were compared to the two min prior to SO (-2 to 0 min). Coordinates of the planes are as in Fig. 3.2.*

Fig. 3.3 reveals the spatial statistical evaluation of the time course of delta activity. Intervals 0-2 (1 min), 4-6 (5 min), and 8-10 min (9 min) were compared to the two min prior to SO -1 min; $P<0.05$, yellow). Recovery exhibited a faster rate of increase with almost 100% of the voxels being different than the 2 minutes before SO already 4-6 min after the transition (Fig. 3.1, right, solid red line), whereas baseline was delayed and converged to the same level as recovery around 8-10 min after SO. The most salient region in both conditions, consistent with Fig. 3.2, is BA 11 in the frontal lobe (Table 3.1), but the changes finally extend to the entire cortex.

Table 3.1 illustrates the three brain regions with the highest delta activity during baseline and recovery of consecutive 2-min intervals, and the brain areas with the largest statistical difference (as measured by *t*-values) between recovery and baseline (activity in recovery > baseline). For example, in the 2-min interval prior to SO of baseline, Brodmann area 4 showed the highest delta activity, Brodmann area 3 the second highest and Brodmann area 6 the third highest level. The corresponding neuroanatomical regions and lobes are also indicated. It is worthy to note that particular Brodmann areas may extend spatially encompassing various neuroanatomical locations, not only a single one. This explains the repetitions occurring in Table 3.1, e.g., BA 11, an important area involved in decision making, planning and reasoning (Rogers et al., 1999), including part of the medial frontal gyrus as well as parts of the superior frontal gyrus in the frontal lobe.

***Table 3.1. (next pages).*** *Brodmann areas and corresponding neuroanatomical regions and lobes showing the 3 strongest activations in the delta band (0.5-5 Hz), for consecutive 2-min intervals in baseline and recovery sleep, and the strongest differences between the two conditions (sorted according to t-values).* $^{*}$ *$P<0.05$, areas that were stronger activated after sleep deprivation than during baseline.*

| **Time interval (min)** | **Baseline** | **Recovery** | **Recovery minus Baseline** |
|---|---|---|---|
| -2-0 | BA 4 (Precentral Gyrus; Frontal Lobe)<br><br>BA 3 (Postcentral Gyrus; Parietal Lobe)<br><br>BA 6 (Precentral Gyrus; Frontal Lobe) | BA 11 (Medial Frontal Gyrus; Frontal Lobe)<br><br>BA 11 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 6 (Superior Frontal Gyrus; Frontal Lobe) | BA 10 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 10 (Middle Frontal Gyrus; Frontal Lobe)<br><br>BA 11 (Middle Frontal Gyrus; Frontal Lobe) |
| 0-2 | BA 3 (Postcentral Gyrus; Parietal Lobe)<br><br>BA 4 (Precentral Gyrus; Frontal Lobe)<br><br>BA 2 (Postcentral Gyrus; Parietal Lobe) | BA 11 (Medial Frontal Gyrus; Frontal Lobe)<br><br>BA 11 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 3 (Postcentral Gyrus; Parietal Lobe) | BA 24* (Anterior Cingulate; Limbic Lobe)<br><br>BA 32* (Anterior Cingulate; Limbic Lobe)<br><br>BA 33* (Anterior Cingulate; Limbic Lobe) |
| 2-4 | BA 3 (Postcentral Gyrus; Parietal Lobe)<br><br>BA 4 (Precentral Gyrus; Frontal Lobe) | BA 11 (Medial Frontal Gyrus; Frontal Lobe)<br><br>BA 11 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 10 (Superior | BA 6* (Middle Frontal Gyrus; Frontal Lobe)<br><br>BA 32* (Cingulate Gyrus; Limbic Lobe)<br><br>BA 24* (Cingulate |

| | | | |
|---|---|---|---|
| | BA 6 (Medial Frontal Gyrus; Frontal Lobe) | Frontal Gyrus; Frontal Lobe) | Gyrus; Limbic Lobe) |
| 4-6 | BA 11 (Medial Frontal Gyrus; Frontal Lobe)<br><br>BA 11 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 10 (Superior Frontal Gyrus; Frontal Lobe) | BA 11 (Medial Frontal Gyrus; Frontal Lobe)<br><br>BA 11 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 10 (Superior Frontal Gyrus; Frontal Lobe) | BA 6* (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 6* (Middle Frontal Gyrus; Frontal Lobe)<br><br>BA 32* (Medial Frontal Gyrus; Frontal Lobe) |
| 6-8 | BA 7 (Precuneus; Parietal Lobe)<br><br>BA 3 (Postcentral Gyrus; Parietal Lobe)<br><br>BA 5 (Postcentral Gyrus; Parietal Lobe) | BA 11 (Medial Frontal Gyrus; Frontal Lobe)<br><br>BA 11 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 10 (Superior Frontal Gyrus; Frontal Lobe) | BA 32* (Cingulate Gyrus; Limbic Lobe)<br><br>BA 24* (Cingulate Gyrus; Limbic Lobe)<br><br>BA 9* (Inferior Frontal Gyrus; Frontal Lobe) |
| 8-10 | BA 11 (Medial Frontal Gyrus; Frontal Lobe)<br><br>BA 11 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 10 (Superior Frontal Gyrus; Frontal Lobe) | BA 11 (Medial Frontal Gyrus; Frontal Lobe)<br><br>BA 11 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 10 (Superior Frontal Gyrus; Frontal Lobe) | BA 32* (Cingulate Gyrus; Limbic Lobe)<br><br>BA 6* (Medial Frontal Gyrus; Frontal Lobe)<br><br>BA 8* (Medial Frontal Gyrus; Frontal Lobe) |

Overall, Table 3.1 supports the notion that a major effect of sleep deprivation was a speeding up of the temporal dynamics of delta activity compared to baseline sleep. We observed that recovery sleep resulted in a temporally accelerated activation of brain areas compared to baseline, showing its highest delta activity in BA 11 and BA 10 (i.e., areas of the prefrontal cortex), stabilizing around 2-4 min after SO. In contrast, baseline dynamics are delayed relative to recovery sleep, as activations in BA 11 and 10 appeared for the first time around 4-6 min after SO, i.e., one 2-min interval later than in recovery. Around 6-8 min after SO, activations in baseline exhibited a shift of its main activated areas to the parietal lobe, but finally returned back to the frontal lobe 8-10 min after SO. Furthermore, BA 11 in recovery even appears activated before SO, although with different regions for second and third Brodmann areas of highest activity, highlighting the role of this prefrontal area in generating delta activity.

### 3.3.2. Temporal evolution of sigma (spindle) activity at sleep onset

We also investigated the temporal evolution of the cortical sources underlying spindle (sigma) activity (12-16 Hz) at the transition into sleep. Highest mean smCSD values in the sigma band did not exceed 0.35 $\mu A^2/(mm^4 \cdot Hz)$, i.e., were typically two orders of magnitude lower than for delta activity (approximately 30 $\mu A^2/(mm^4 \cdot Hz)$; Figs. 3.1 and 3.4, left panels, blue and red lines).

Mean sigma activity gradually increased after SO in both conditions (Fig. 3.4, left, blue and red line) and peaked around 7 min after SO in baseline and at 3 min in recovery. The largest increase occurred in the first 2-min interval after SO. Activations did not differ between conditions, i.e. hardly any voxels differed between the two conditions (Fig. 3.4, left, green line).

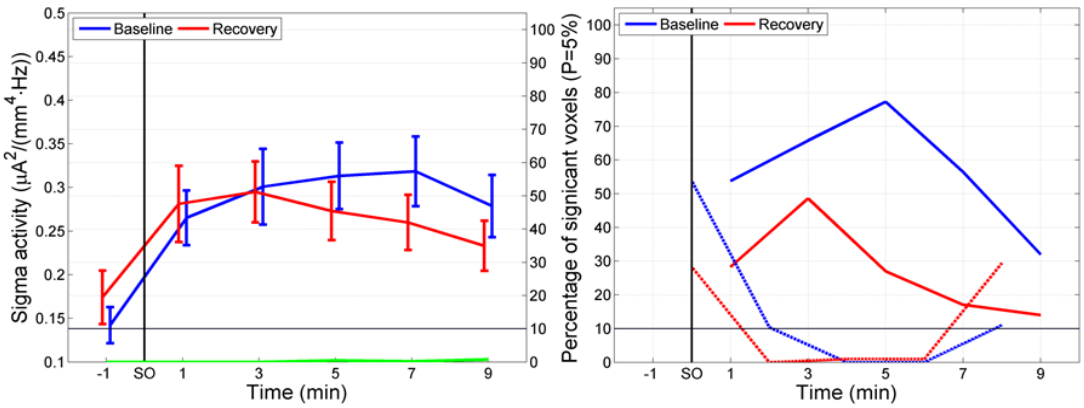


***Fig. 3.4.*** *Left panel: Temporal evolution of sigma activity (mean squared magnitude of current source density in the sigma band (12-16 Hz) [μA²/(mm⁴·Hz)]; averaged across all voxels; left y-axis) and statistical evaluation (percentage of significantly different voxels between conditions; right y-axis) at the transition to sleep (sleep onset, SO). Right panel: Percentage of voxels significantly different from the 2-min interval prior to SO (solid lines) and percentage of voxels significantly different between consecutive 2-min intervals (dashed lines).*

The percentage of voxels differing from the 2 min prior to SO peaked at 5 min in baseline and at 3 min in recovery and was substantially larger in baseline (Fig. 3.4, right, solid lines). In summary, in baseline sleep larger brain areas produced spindle activity and, furthermore, were able to sustain it for a longer time. Consecutive 2-min intervals only showed a difference in the percentage of voxels activated at the transition and in the last 2-min interval in both conditions (Fig. 3.4, right, dashed lines), corresponding to an increase in sigma activity after SO (0-2 min) and a decrease in the last 2-min interval (8-10 min).

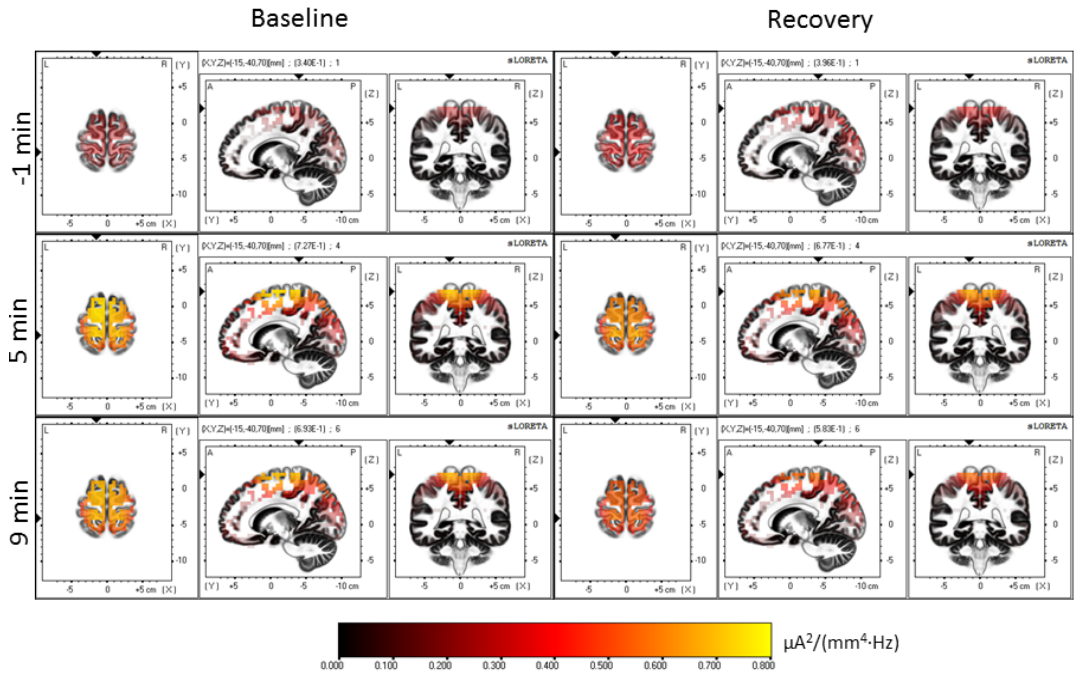


***Fig. 3.5.*** *LORETA source estimation in the sigma band (squared magnitude of current source density [μA²/(mm⁴·Hz)] in the sigma band (12-16 Hz). Left: baseline, right: recovery sleep. Images represent average values of the two minutes before SO (-1 min), 4-6 min (5 min) and 8-10 min (9 min) after SO. The plane coordinates were chosen to best match the most active brain regions (-15, -40, 70 in MNI coordinates [mm]).*

Fig. 3.5 depicts the smCSD maps illustrating the underlying generators of spindle activity. They highlight particularly the parietal lobe, both in baseline and recovery, as the most prominent region for the generation of cortical spindles. However, there were some additional contributions, although more diffuse, resulting from posterior parts of the frontal lobe (such as the postcentral gyrus) and, towards the back of the brain, extending marginally into the cuneus and the lingual gyrus of the occipital cortex (Table 3.2).

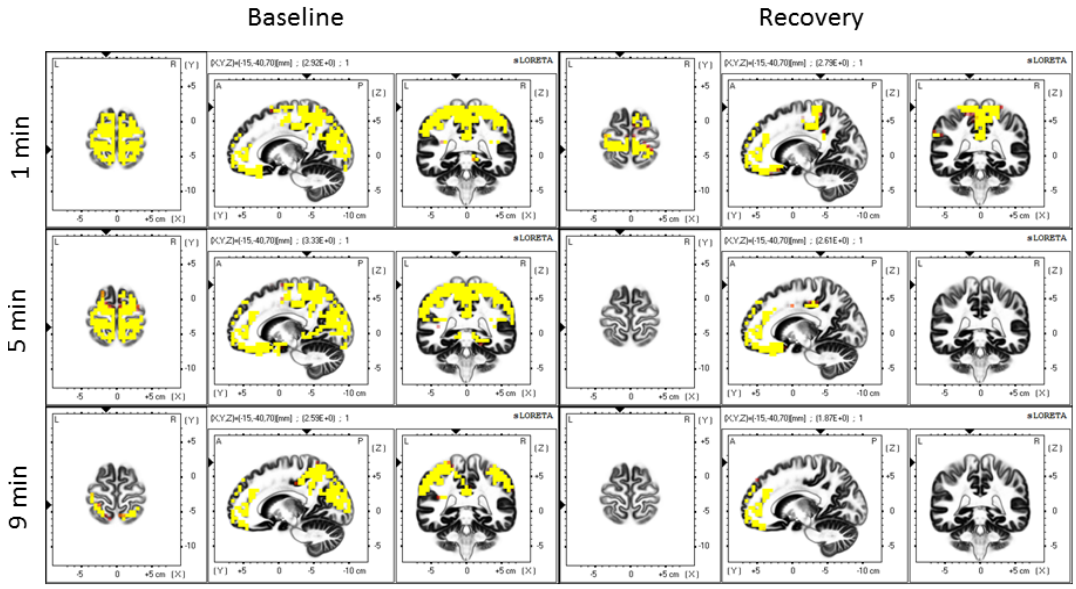


***Fig. 3.6.*** *Statistical evaluation of the temporal evolution of sigma activity (12-16 Hz). Left: baseline, right: recovery sleep. Yellow areas represent voxels that showed higher sigma activity after SO (t-values at P<0.05). Minutes 0-2 (1 min), 4-6 (5 min) and 8-10 (9 min) were compared to the two minutes prior to SO (-2 to 0 min). Coordinates of the planes are as in Fig. 3.5.*

The temporal evolution of spatial aspects of sigma activity is illustrated in Fig. 3.6. Minutes 0-2 (1 min), 4-6 (5 min) and 8-10 (9 min) after SO were compared to the two minutes prior to SO (-2 to 0 min). The baseline condition (Fig. 3.6, left) revealed in general more widespread statistical changes across the brain after SO than during recovery, i.e. covering larger areas. The capacity to generate spindles was reduced by sleep deprivation (Fig. 3.6, right). It is worthy to note the symmetry between the left and right hemispheres in their statistical topographical properties, as shown by the coronal plane section (right map of each row). The most consistent regions showing changes in sigma activity appeared along the parietal lobe and its cortical surroundings, with some contributions of frontal areas (Table 3.2).

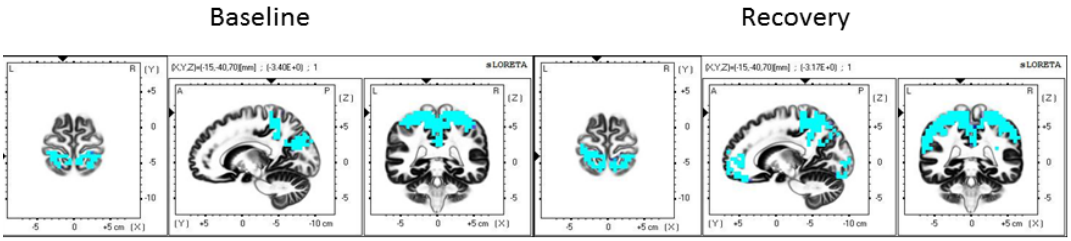


***Fig. 3.7.*** *Statistical evaluation of the temporal evolution of sigma activity (12-16 Hz). Left: baseline, right: recovery sleep. Blue areas represent voxels that decreased in activity in the statistical contrast (t-values at P<0.05). Minutes 8-10 (last interval) were compared to the previous two min, 6-8 min after SO. Coordinates of the planes are as in Fig. 3.5.*

A further aspect of the temporal evolution, comparing consecutive 2-min intervals, is depicted in Fig. 3.7 comparing 8-10 min after SO with the preceding 2 min. The first 2 min after SO were already illustrated in Fig. 3.6. These are the only two intervals showing at least 10% of all voxels changed (Fig. 3.4, right, dashed curves). The first 2 min after SO showed increased sigma activity (Fig. 3.6, top row), the last 2 min a decrease (Fig. 3.7). Fig. 3.7 is also consistent with Figs. 3.5 and 3.6, showing again that the areas playing a predominant role for the sigma activity at SO were mainly located in the parietal lobe.

***Table 3.2. (next pages)***. *Brodmann areas and corresponding neuroanatomical regions and lobes showing the 3 strongest activations in the sigma band (12-16 Hz), for consecutive 2-min intervals in baseline and recovery sleep, and the strongest differences between the two conditions (sorted according to t-values).* * *P<0.05, areas that were less activated after sleep deprivation than during baseline. Significant differences appeared only in the last two 2-min intervals.*

| **Time interval (min)** | **Baseline** | **Recovery** | **Recovery minus Baseline** |
|---|---|---|---|
| -2-0 | BA 4 (Precentral Gyrus; Frontal Lobe)<br><br>BA 3 (Postcentral Gyrus; Parietal Lobe)<br><br>BA 4 (Postcentral Gyrus; Frontal Lobe) | BA 7 (Precuneus; Parietal Lobe)<br><br>BA 19 (Cuneus; Occipital Lobe)<br><br>BA 19 (Precuneus; Parietal Lobe) | BA 19 (Fusiform Gyrus; Occipital Lobe)<br><br>BA 18 (Lingual Gyrus; Occipital Lobe)<br><br>BA 19 (Lingual Gyrus; Occipital Lobe) |
| 0-2 | BA 3 (Postcentral Gyrus; Parietal Lobe)<br><br>BA 4 (Precentral Gyrus; Frontal Lobe)<br><br>BA 6 (Precentral Gyrus; Frontal Lobe) | BA 4 (Precentral Gyrus; Frontal Lobe)<br><br>BA 6 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 6 (Medial Frontal Gyrus; Frontal Lobe) | BA 9 (Middle Frontal Gyrus; Frontal Lobe)<br><br>BA 9 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 10 (Middle Frontal Gyrus; Frontal Lobe) |
| 2-4 | BA 4 (Precentral Gyrus; Frontal Lobe)<br><br>BA 3 (Postcentral Gyrus; Parietal Lobe)<br><br>BA 6 (Superior Frontal Gyrus; Frontal Lobe) | BA 3 (Postcentral Gyrus; Parietal Lobe)<br><br>BA 4 (Precentral Gyrus; Frontal Lobe)<br><br>BA 4 (Paracentral Lobe; Parietal Lobe) | BA 18 (Middle Occipital Gyrus; Occipital Lobe)<br><br>BA 17 (Cuneus; Occipital Lobe)<br><br>BA 19 (Middle Occipital Gyrus; Occipital Lobe) |

| | | | |
|---|---|---|---|
| 4-6 | BA 6 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 6 (Medial Frontal Gyrus; Frontal Lobe)<br><br>BA 6 (Precentral Gyrus; Frontal Lobe) | BA 3 (Postcentral Gyrus; Parietal Lobe)<br><br>BA 4 (Precentral Gyrus; Frontal Lobe)<br><br>BA 4 (Paracentral Lobe; Parietal Lobe) | BA 47* (Inferior Frontal Gyrus; Frontal Lobe)<br><br>BA 11* (Inferior Frontal Gyrus; Frontal Lobe)<br><br>BA 11* (Middle Frontal Gyrus; Frontal Lobe) |
| 6-8 | BA 3 (Postcentral Gyrus; Parietal Lobe)<br><br>BA 4 (Precentral Gyrus; Frontal Lobe)<br><br>BA 4 (Paracentral Lobe; Parietal Lobe) | BA 3 (Postcentral Gyrus; Parietal Lobe)<br><br>BA 4 (Precentral Gyrus; Frontal Lobe)<br><br>BA 4 (Postcentral Gyrus; Frontal Lobe) | BA 31* (Cingulate Gyrus; Limbic Lobe)<br><br>BA 23* (Posterior Cingulate; Limbic Lobe)<br><br>BA 30* (Posterior Cingulate; Limbic Lobe) |
| 8-10 | BA 6 (Superior Frontal Gyrus; Frontal Lobe)<br><br>BA 6 (Medial Frontal Gyrus; Frontal Lobe)<br><br>BA 3 (Postcentral Gyrus; Parietal Lobe) | BA 3 (Postcentral Gyrus; Parietal Lobe)<br><br>BA 4 (Precentral Gyrus; Frontal Lobe)<br><br>BA 4 (Postcentral Gyrus; Frontal Lobe) | BA 6* (Precentral Gyrus; Frontal Lobe)<br><br>BA 4* (Precentral Gyrus; Frontal Lobe)<br><br>BA 18* (Cuneus; Occipital Lobe) |

Table 3.2 lists the neuroanatomical location of the most salient regions appearing in Figs. 3.5, 3.6 and 3.7. Thus, both for baseline and recovery, the most relevant Brodmann areas were, rated in order of the strength of their activity levels: BA 3, 4 and 6. BA 3 and 4 are located in between the superior parietal lobe and the posterior part of the frontal lobe, near the paracentral lobe (part of BA 4) and the central sulcus; BA 6 is on the posterior edge of the frontal lobe. Table 3.2 also reveals that for the second 2-min interval after SO (2-4 min after SO) and fourth interval (6-8 min after SO), there was an exact match of the three most activated Brodmann areas in baseline and recovery. Additionally, by counting the lobes appearing in each column, the table suggests that, although sigma activity is principally originating from centro-parietal regions, for baseline, there is a small preference towards more anterior areas (particularly to BA 6, on the edge of the frontal lobe, with seven repetitions in the second column), whereas for recovery, the preference goes more towards posterior areas (particularly to BA 4, encompassing the central sulcus, with eight repetitions in the third column).

### 3.3.3. Temporal evolution of alpha activity at sleep onset

Alpha activity (8-12 Hz) exhibits a progressive increase with time reaching a plateau during the last five minutes (Fig. S3.1, blue and red lines). This time course is observed in both conditions (particularly in parieto-occipital areas), but in recovery the increase was faster and a higher asymptotic value than in baseline was reached. Supplementary Fig. S3.2 indicates that strongest alpha activity was observed in BA 7, the precuneus (on the edge between the superior parietal and occipital lobes). The prominent role of the precuneus in the alpha band is consistent with the sleep onset process, since it is an area previously linked to visual imagery (i.e., it might be related to hypnagogic hallucinations) and critical for conscious information processing (Vogt and Laureys, 2005). More detailed information and descriptions can be found in Supplementary Material.

### 3.3.4. Temporal evolution of beta activity at sleep onset

Beta activity (16-24 Hz) showed a decrease with time in both conditions, but with a faster decline in recovery (Fig. S3.4, blue and red line). However, less than 10% of the voxels differed between the conditions (Fig. S3.4, green line). The most salient regions (based on maximum smCSD values) were in the occipital and parietal lobes. Supplementary Figs. S3.5 and S3.6 reveal topographically the decrease of beta activity in grey matter voxels and indicate that for the recovery condition beta activity dissipated quicker. As for alpha activity, BA 7 (the precuneus) was also involved in the generation of beta activity. In contrast, beta activity was decreasing while alpha activity was increasing. Other major areas involved in the generation of beta activity were BA 6 and 4. More detailed information is provided in Supplementary Material.

### 3.3.5. Temporal evolution of theta activity at sleep onset

Similar to alpha activity, theta activity (5-8 Hz) increased in both conditions until a plateau was reached (Fig. S3.7, blue and red line). The plateau levels were higher in recovery than baseline. Theta activity did not differ between conditions prior to SO (Figs. S3.7, green line and S3.8, upper row), whereas there was a considerable statistical separation after SO (Figs. S3.7 to S3.9), indicating that the maximal statistical separation between conditions happened at 5 min after SO (with more than 60% of the voxels being significantly different). The most salient areas in the smCSD maps were again the precuneus (BA 7), in the superior parietal lobe, followed by the cuneus (BA 19), in the occipital lobe. For more detailed information, see Supplementary Material.

## 3.4. Discussion

We investigated the strength and distribution of the intracranial sources of activity of different frequency bands at the transition into sleep (first ten minutes) and how they are affected by increased sleep pressure (total sleep deprivation). We applied LORETA (Pascual-Marqui et al., 2002) to estimate the underlying sources of brain electrical activity. Changes in delta and alpha activity at sleep onset involved large areas of the cortex, whereas activity in the beta range was restricted to small cortical areas. Increased sleep pressure due to sleep deprivation altered the dynamics in recovery sleep leading to kind of a "fast-forward", i.e., accelerated version of the baseline dynamics appearing in most frequency bands.

### 3.4.1. Delta activity

Delta activity averaged across all cortical gray matter voxels showed an exponential increase at the beginning of sleep (Fig. 3.1, left, blue and red lines) which was faster after sleep deprivation. This finding is in line with early studies examining the buildup of SWA of a single EEG derivation (C3A2; (Dijk et al., 1990)). Siclari et al. (Siclari et al., 2014) demonstrated a similar time course of slow wave density and hypothesized that the observed temporal evolution resulted from two distinct synchronization processes, first dominated by "Type I" slow waves (mediated by subcortical arousal promoting structures in the pons), followed by "Type II" slow waves (originating from cortico-cortical synchronization processes).

Previous studies have shown that anterior areas are the first to exhibit slow waves activity, and the onset of these waves is accelerated in recovery sleep (De Gennaro et al., 2005b, Ferrara and De Gennaro, 2011). Already in 1949, Brazier (Brazier, 1949) reported a shift of focus from the occipital to frontal regions when transitioning from waking to sleep. Gradually more brain regions showed increased slow wave activity compared to the pre-sleep interval reaching 100% after 9 min in baseline and almost 100% at 5 min in recovery (Fig. 3.1). After 3 min more than 70% of the voxels differed in intensity between the conditions reaching 100% in the last 2-min interval. Brain regions that

got gradually involved (as assessed by showing the largest smCSD values) were BA 3 (postcentral gyrus in the parietal lobe), BA 6 (premotor cortex and supplementary motor cortex in the frontal lobe) and BA 10 and 11 (frontopolar cortex in the frontal lobe) (Fig. 3.2 and 3.3; Table 3.1). Siclari et al. (Siclari et al., 2014) reported that the brain regions predominantly involved in the rise of delta activity were located in the anterior medial prefrontal cortex, primary cortex and posteromedial parietal cortex.

As revealed by the topographic analysis and illustrated in smCSD maps, the homeostatic effects were accentuated over the frontal lobe. These regions experience a higher computational load due to daily activities, such as decision making, executive functions and emotional regulation (Rogers et al., 1999, Nir et al., 2011, Ikeda et al., 2015). The study of Siclari et al. (Siclari et al., 2014) shows agreement with our own results. Power spectral density increased progressively along the 10 intervals covered (nearly 10 min), being more prominent in the frontal lobe and central areas of the parietal lobe. The most frequent origins of slow waves at the beginning of SO corresponded to the primary motor and sensory cortices (extending also into dorsolateral prefrontal cortex), the inferior frontal gyrus and the posteromedial parietal cortex (posterior cingulate and precuneus). Cortical involvement maps (the amount of current produced by different cortical areas) also highlighted the major role exerted by the frontal lobe at the beginning of the SO transition. Thus, the areas reported by Siclari et al. (Siclari et al., 2014) are concordant with our topographical results (Figs. 3.2 and 3.3 and Table 3.1). Towards the second half of the time interval investigated, the origin of slow waves spread evenly throughout the cortex, which is also in correspondence with the spatially homogenous maps that we observed in the last 2-min interval.

De Gennaro et al. investigated the behavior of power spectral density along the SO transition (De Gennaro et al., 2000b, De Gennaro et al., 2001b). Based on the scalp EEG, they found that in a five min interval before SO (defined as the first appearance of a spindle or K-complex), the low frequency range (<7 Hz) was characterized by an anterior-posterior gradient, in which central, frontal and frontopolar areas stand

out the most. Parietal areas also showed more prominence than occipital areas. On the other hand, in a five min interval after SO, power was significantly larger in fronto-central scalp locations. They argued that the prevalence of EEG power in frontal and central areas indicate that these areas are the first to synchronize its oscillatory activity during the SO process. These results are consistent with what we observed through source localization. Another study by the same authors confirmed and extended the previous results (De Gennaro et al., 2001a). In the low frequency range (1-6 Hz), a linear increase of power spectral density with time was observed. In addition, power was significantly higher after stage 2 onset than during stage 1. A polynomial fit confirmed the accelerated dynamics of the delta rise after the SO transition, matching what we observed.

### 3.4.2. Sigma activity

Sigma activity (12-16 Hz) was found to follow an inverted U-shape along the SO transition, peaking at a different time interval depending on the level of sleep pressure (quicker in recovery; Fig. 3.4) and showing, at the beginning of the SO transition, a direct relationship with delta activity, but an inverse relationship with delta at the end of the window. This type of inverted U-shape behavior has been previously observed in single channel EEG studies (Aeschbach and Borbély, 1993, Aeschbach et al., 1997). In addition, Siclari et al. reported a similar trend in the rise of spindle density and the percentage of scalp involvement in spindle activity across channels (Siclari et al., 2014). During recovery, spindle activity was diminished compared to baseline in accordance with previous reports (Borbély et al., 1981, De Gennaro et al., 2001a).

Commonly, fast and slow spindles are dissociated (De Gennaro and Ferrara, 2003). We focused on fast spindle activity (12-16 Hz). It is, however, difficult to distinguish slow spindles from alpha activity, in particular alpha activity occurring in NREM sleep (Scheuler et al., 1988, Scheuler et al., 1993, Olbrich and Achermann, 2005). Additionally, we investigated the changes in the alpha band (8-12 Hz), which comprises both NREM alpha and slow spindle activity (see below).

De Gennaro et al. investigated the effect of two nights of selective slow wave deprivation on spindle density (12-15 Hz) (De Gennaro and Ferrara, 2003). An inverse relationship between delta and sigma activity, such that an enhancement of delta activity accompanied by a decrease of sigma activity was observed. A simple topographical analyses of sleep spindle density of the channels studied revealed spatial heterogeneity of spindle occurrence, where centro-parietal areas revealed stronger spindle activity, followed by the frontal lobe, and lastly the occipital cortex. These results are corroborated by our own analyses (Fig. 3.4 for the mean evolution of spindle activity; Figs. 3.5 and 3.6 as well as Table 3.2 for topographical aspects).

Also the results of spindle activity in the study performed by Siclari et al. (Siclari et al., 2014), were largely congruent with our own analyses. The spindle density (number of spindles per minute) underwent an initial increase until reaching a maximum (appearing earlier and more pronounced on anterior and central regions), which then experienced a moderate decrease during the last minutes of the nearly 10-min window investigated. Furthermore, the authors distinguished between slow frontal and fast centro-parietal spindles, the latter corresponding to our sigma activity (12-16 Hz). The most frequent origin of spindles assessed by Siclari et al. (Siclari et al., 2014) was located in the anterior cingulate cortex and the dorsolateral and medial prefrontal cortex for frontal spindles, whereas it was located in the precuneus and posterior cingulate cortex for centro-parietal spindles (all these areas also appeared statistically involved in our study; Figs. 3.5 and 3.6, Table 3.2, particularly in the baseline condition).

Del Felice et al. investigated the cortical sources of sleep spindles using LORETA (Del Felice et al., 2014). They localized slow (10-12 Hz) and fast (12-14 Hz) spindles. Fast spindles appeared more scattered throughout the brain, including the medial and inferior frontal gyrus, inferior parietal lobule, middle temporal gyrus and lingual gyrus in the occipital cortex. The reported brain areas are comparable with our own results. Slow spindles were mostly localized along the frontal lobe (in particular, medial and inferior frontal gyri).

### 3.4.3. Other frequency bands

In the alpha band (8-12 Hz), activity gradually increased along the SO transition until reaching a steady state involving 50 to 80% of the cortex. The observed rise was particularly pronounced on the occipital cortex and parts of the parietal lobe and during recovery sleep a faster increase was present. The behavior of the theta band (4-8 Hz) was very similar to the alpha band.

Beta activity gradually diminished. This behavior has also been observed in other studies (De Gennaro et al., 2001a, De Gennaro et al., 2005b). The exponential decay was quicker in the recovery, hence, reflecting the higher homeostatic pressure in this condition. Areas displaying the most prominent activity involved parieto-occipital regions with some frontal contributions.

Park et al. also investigated the SO transition using standardized LORETA (30 EEG derivations) (Park et al., 2015). Sleep onset, was defined as the first occurrence of stage 1, characterized by increased delta and theta activity in the occipital cortex, while low alpha (8-10 Hz) activity decreased, early in stage 1. On the other hand, theta power increased in parietal structures (also observed in our study in Supplementary Fig. S3.8). Subsequently, alpha activity decreased throughout most brain areas, while theta activity increased along parieto-occipital regions. Later, towards the end of stage 1, beta activity decreased mainly in the frontal lobe, followed by some parietal, temporal and limbic (cingulate gyri, parahippocampal gyrus and posterior cingulate) structures. The most significant increase of theta activity appeared in the anterior cingulate cortex and adjacent structures of the frontal pole (see last row of Supplementary Fig. S3.9 showing the same structures (yellow voxels, indicating statistical significance), and moreover, indicating that this change occurred around 9 min after SO).

## 3.5. Limitations

Some limitations of the current analysis are important to note. For one, we could not address subcortical contributions, as the solution space is limited to the cortex (Pascual-Marqui et al., 1994, Pascual-Marqui, 2002). Furthermore, subcortical components are often too weak to contribute directly to the EEG measured at the scalp level (Pascual-Marqui, 2002, Pascual-Marqui et al., 2011). We considered applying a new approach (Krishnaswamy et al., 2017) which claims that subcortical sources can be distinguished from cortical ones when the underlying cortical activity is sparse (i.e., limited to a subset of cortical regions). Yet, we decided against this approach, given the wide spread cortical activity of slow waves in sleep.

## 3.6. Conclusions

LORETA proved to be a valuable tool to reveal the cortical sources underlying brain oscillatory activity at the transition into sleep. Classical EEG frequency bands exhibited location dependent brain dynamics during the SO transition. In the delta band, the resulting smCSD maps rendered BA 11 as the most relevant brain region associated with delta activity, both in baseline and recovery (the latter condition reaching much stronger intensity levels). Additionally, the statistical divergence between baseline and recovery increased steadily during SO. The dynamics of spindle activity resembled in both conditions an inverted U-shape. The maximum of activity was shifted closer to SO in recovery and the capacity to sustain spindle activity was impaired.

## Acknowledgements

This project was supported by the Swiss National Science Foundation, grant 32003B_146643. We thank Drs. Leila Tarokh and Pascal Faber for comments on the manuscript.

## 3.7. Supplementary Material

In the Supplementary Material, we present next to some methodological aspects the results for further frequency bands (in alphabetical order): alpha (8-12 Hz), beta (16-24 Hz) and theta (4-8 Hz). In the main paper, the focus was on the delta band (slow-wave activity, related to sleep homeostasis) and the sigma band (spindle activity). In this way, we complement the main analyses and render a more complete picture of the dynamics and behavior of the brain during the SO transition.

### 3.7.1. Supplementary Methods

Classical electrodynamic theory establishes a forward equation that, in the context of the EEG, links current source density (CSD) inside the brain to the voltage potential differences measured at the scalp (Lantz et al., 1997, Pascual-Marqui et al., 2014a, Pascual-Marqui et al., 2014b). The forward equations relates to the forward problem, which mathematically is a well-posed problem, and consequently, has a well-defined (unique) solution, although typically is solved with numerical methods in the lack of a closed analytical expression (Pascual-Marqui et al., 2014a). However, the inverse problem, that aims to obtain the cortical activity from the recorded time series of the EEG, constitutes an ill-posed problem, which is known since the XIX century to have no unique solution, but potentially an infinite number of solutions (Vanoosterom, 1991, Astolfi et al., 2007).

In order to circumvent the ill-posed nature of the inverse problem for CSD estimation, some additional constraints are required to be imposed that limit the infinite range of solutions to a single one, hence, suited constraints need to be chosen (Lantz et al., 1997, Pascual-Marqui et al., 2002). In this study, “exact low resolution electromagnetic tomography” (eLORETA, or shortly, LORETA) (Pascual-Marqui et al., 2014a) was adopted. The LORETA solution is characterized by imposing maximal smoothness in the solution of the underlying cortical generators (for a given voxel, a surrounding small

sphere would have, on average, minimum difference), and has the virtue of zero localization error, but on the other hand, exhibits relatively poor spatial resolution (Pascual-Marqui, 2002). LORETA images are cortical spatial maps of squared magnitude of current source density (smCSD) where each voxel has an associated single value, therefore, forming a scalar map. Given the fact that the dipole currents corresponding to the generators are vectors with three components (as they are embedded in a 3-d space), a brief explanatory clarification needs to be provided for this apparent inconsistency in the dimensionality. Although in rigorous terms the solution is indeed a vector field, a crucial constraint is imposed to the time series of the generators, given the fact that pyramidal neurons (the ones that mostly contribute to the EEG) are oriented along a spatially fixed direction, orthogonal to the cortical surface of the brain (Babiloni et al., 2005, Astolfi et al., 2007). As a consequence, the generators in each voxel can be expressed, with no significant loss of information, just with the maximum covariance eigenvector, which yields a single degree of freedom, hence, representing a cortical scalar map for univariate time series associated with smCSD (Pascual-Marqui et al., 2014a).

### 3.7.2. Supplementary Results

#### 3.7.2.1. Temporal evolution of alpha activity at sleep onset

Mean alpha activity (8-12 Hz) showed a saturating increase in the first 10 min after SO in both conditions, reaching a plateau at 5 min after the SO transition. Approximately 15% of the voxels differed between the baseline and recovery conditions after SO (Fig. S3.1, left, green line).

The percentage of voxels differing from the 2 min prior to SO was larger in recovery than in baseline. Approximately 35% of voxels differed already in the first 2-min interval after SO (Fig. S3.1, right, solid lines) whereas in baseline differences emerged at 3 min after SO. Later, 80% of the voxels revealed higher activity than prior to SO in recovery sleep and only about 50% of the voxels in baseline. Comparing consecutive 2-min intervals, in baseline only the 2$^{nd}$ interval (3 min) differed from the 1$^{st}$ one (1 min; Fig. S3.1, right, dashed lines at the bottom). During recovery, interval 1 (1 min) and 2 (3 min) differed from the previous ones.

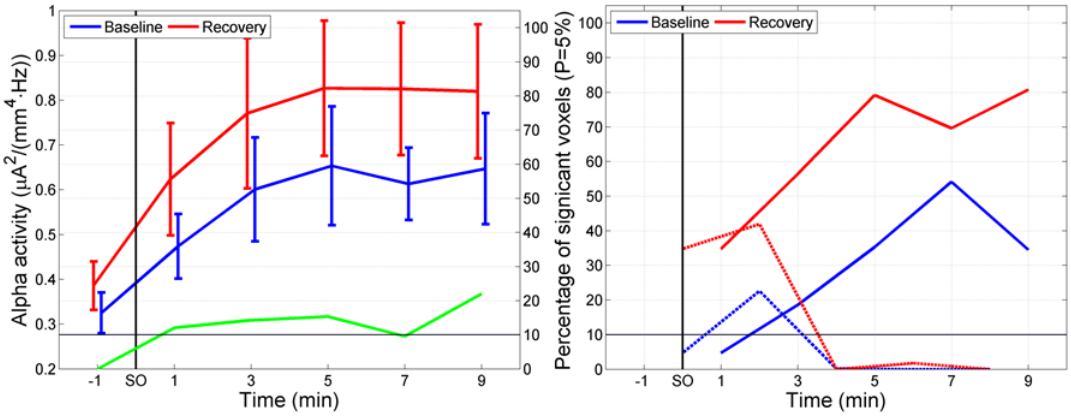


***Fig. S3.1.*** *Left panel: Temporal evolution of alpha activity (mean squared magnitude of current source density in the alpha band 8-12 Hz; [μA²/(mm$^4$·Hz)]; averaged across all voxels; left y-axis) and statistical evaluation (percentage of significantly different voxels between conditions; right y-axis) at the transition to sleep (sleep onset, SO). Blue line: mean alpha activity in baseline; red line: mean alpha activity in recovery; green line: percentage of voxels significantly different between recovery and baseline. Right panel: Percentage of voxels significantly different from the 2-min interval prior to SO (solid lines) and percentage of voxels significantly different between consecutive 2-min intervals (dashed lines). Blue lines: baseline; red lines recovery.*

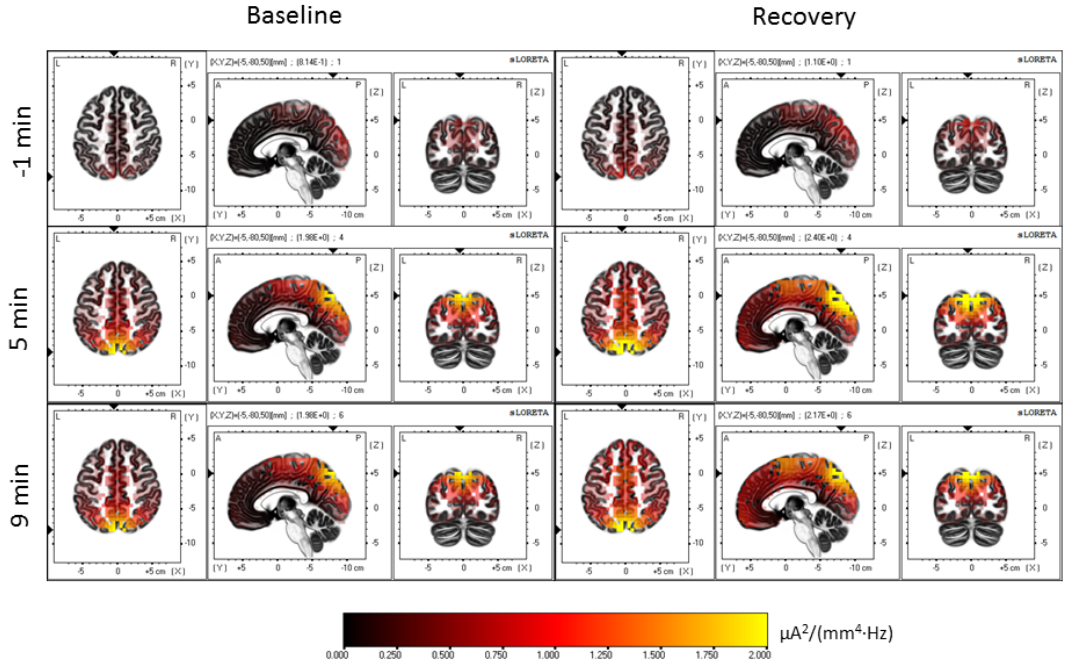


***Fig. S3.2.*** *LORETA source estimation of alpha activity (squared magnitude of current source density [µA²/(mm$^4$·Hz)]) in the alpha band (8-12 Hz). Left: baseline, right: recovery sleep. Images represent average values of two minutes before SO (-1 min), 4-6 min (5 min) and 8-10 min (9 min) after SO. The plane coordinates were chosen to best match the most active brain regions (-5, -80, 50; MNI coordinates [mm]).*

Fig. S3.2 illustrates the gradual spatial buildup of alpha activity (two minutes before SO (-1 min), 4-6 (5 min) and 8-10 min (9 min) after SO), both, in baseline (left) and recovery sleep (right). The neuroanatomical region of highest alpha activity in both conditions, and for all considered time intervals, was BA 7, the precuneus (mainly located in the superior parietal lobe, but also stretching to some parts of the occipital lobe). The precuneus is an area known to play a major role in consciousness and visuospatial imagery (Vogt and Laureys, 2005). However, in recovery sleep, alpha activity is not only more intense in absolute terms, but also distributed across a much broader set of cortical areas, even encompassing the frontal lobe during the last minutes (8-10 min).

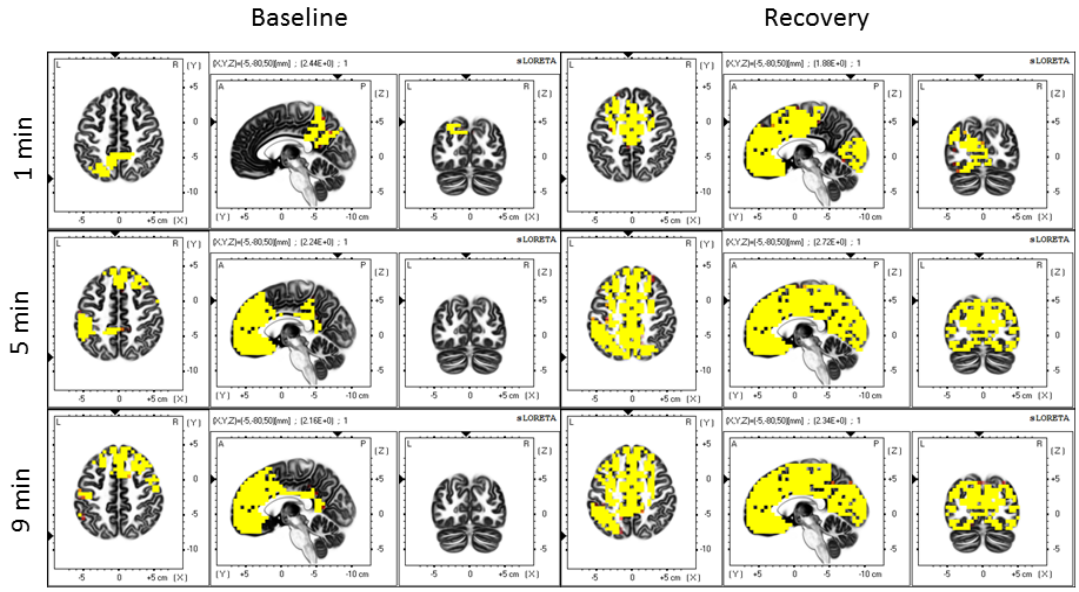


***Fig. S3.3.*** *Statistical evaluation of the temporal evolution of alpha activity (8-12 Hz). Left: baseline, right: recovery sleep. Yellow areas represent voxels that showed higher alpha activity after SO (t-values at P<0.05). Minutes 0-2 (1 min), 4-6 (5 min) and 8-10 (9 min) were compared to the two min prior to SO (-2 to 0 min). Coordinates of the planes are as in Fig. S3.2.*

Fig. S3.3 illustrates the statistical comparisons of minutes 0-2 (1 min), 4-6 (5 min) and 8-10 (9 min) with to the two min prior to SO (-2 to 0 min). Immediately after SO (upper row), areas of the occipital cortex started to show increased alpha activity, both during baseline and recovery. However, for recovery, additional areas comprising the whole frontal lobe and parts of the parietal cortex also revealed a rise in alpha activity. With increasing time into sleep, the recruitment of the frontal lobe also in baseline became evident, whereas in recovery the changes covered large parts of the cortex (approximately 80%, Fig S3.1, right, solid red line). In both conditions, some inter-hemispheric asymmetries can be appreciated.

### 3.7.2.2. Temporal evolution of beta activity at sleep onset

Mean beta activity (16-24 Hz) decreased as a function of time (Fig. S3.4, blue and red lines) in both conditions. However, less than 10% of voxels differed between the conditions (green line). The decline could be approximated by an exponential function (~exp($\alpha \cdot t$); $t$, time), in particular during recovery. Excluding the values prior to SO, the fit revealed $\alpha$=-0.045±0.017 min$^{-1}$ for baseline and $\alpha$=-0.071±0.026 min$^{-1}$ for recovery (95% CI). At 7 and 9 min after SO, 20 and 30% of the voxels differed from the 2 min prior to SO in baseline and recovery, respectively.

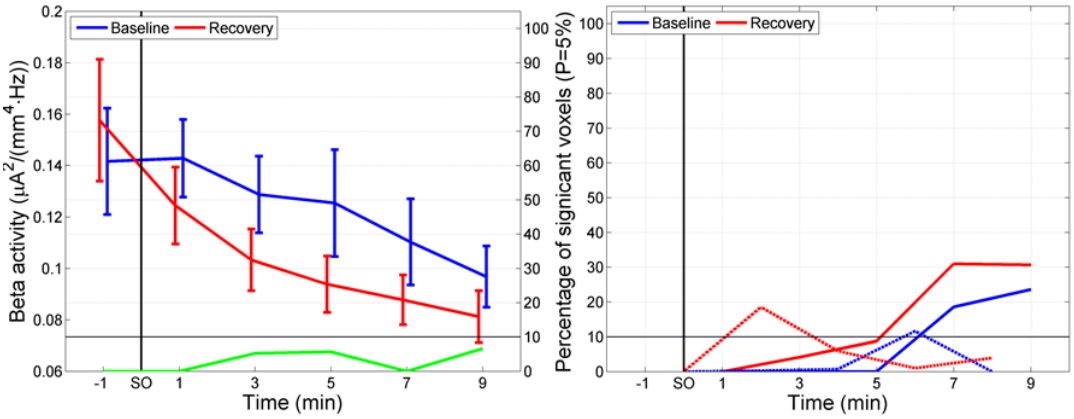


***Fig. S3.4.*** *Left panel: Temporal evolution of beta activity (mean squared magnitude of current source density in the beta band (16-24 Hz) [μA²/(mm⁴·Hz)]; averaged across all voxels; left y-axis) and statistical evaluation (percentage of significantly different voxels between conditions; right y-axis) at the transition to sleep (sleep onset, SO). Right panel: Percentage of voxels significantly different from the 2-min interval prior to SO (solid lines) and percentage of voxels significantly different between consecutive 2-min intervals (dashed lines).*

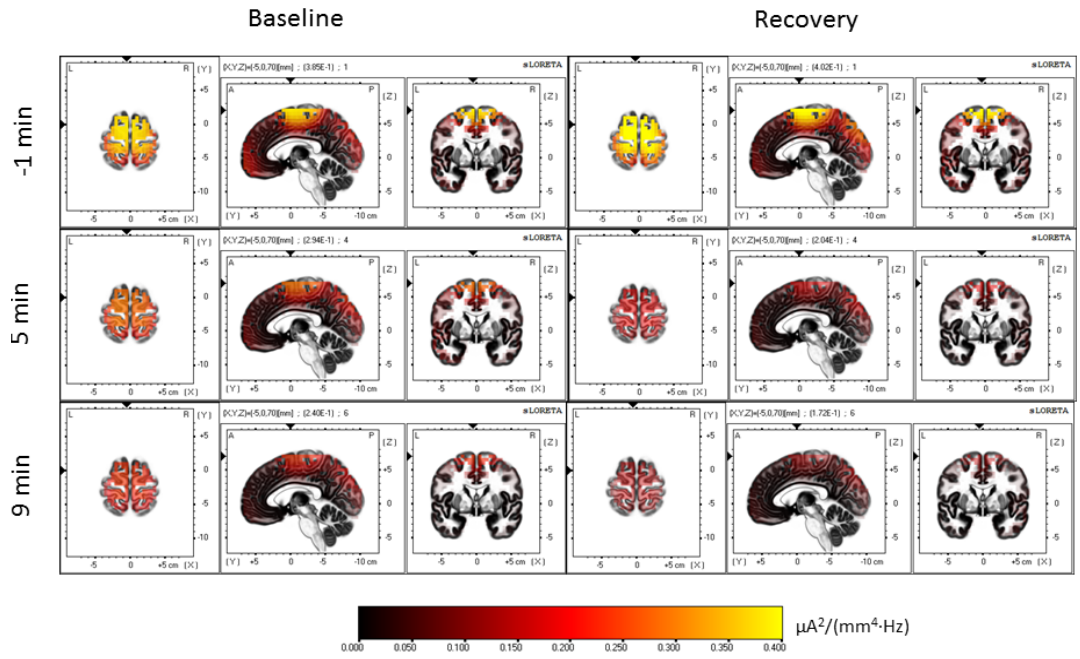


***Fig. S3.5.*** *LORETA source estimation in the beta band (squared magnitude of current source density [μA²/(mm⁴·Hz)] in the beta band (16-24 Hz)). Left: baseline, right: recovery sleep. Images represent average values of two minutes before SO (-1 min), 4-6 min (5 min) and 8-10 min (9 min) after SO. The plane coordinates were chosen to best match the most active brain regions (-5, 0, 70 in MNI coordinates [mm]).*

Fig. S3.5 represents smCSD maps of beta activity two minutes before SO (-1 min), 4-6 (5 min) and 8-10 min (9 min) after SO, both in baseline (left) and recovery sleep (right). The most relevant brain areas, sorted according to their smCSD values (which decreased with time), comprised firstly, the precuneus (on the edge of the parietal lobe, BA 7) that exhibited the highest values in all intervals, secondly, the superior frontal gyrus (in the frontal lobe, BA 6), and thirdly, the paracentral gyrus (located between the frontal and parietal lobes, BA 4).

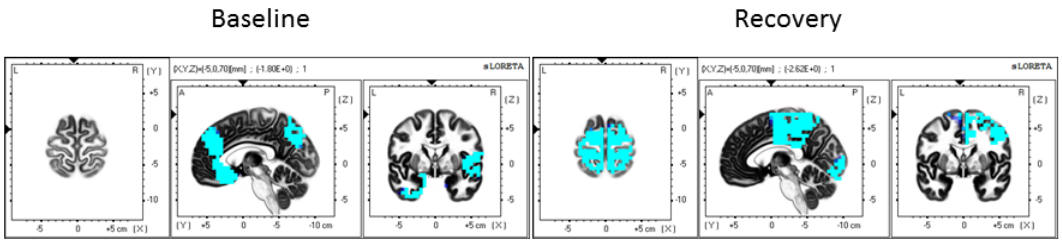


***Fig. S3.6.*** *Statistical evaluation of the temporal evolution of beta activity (16-24 Hz). Left: baseline, right: recovery sleep. Blue areas represent voxels that showed lower beta activity after SO (t-values at P<0.05). Minutes 8-10 were compared to the two min prior to SO (-2 to 0 min). Coordinates of the planes are as in Fig. S3.5.*

Fig. S3.6 illustrates the statistical comparison of min 8-10 with to the two min prior to SO where more than 10% of the voxels differed. In baseline, the brain regions that exhibited the largest change comprised the ventromedial prefrontal cortex and the dorsomedial prefrontal cortex in the frontal lobe (BA 10, 11 and 12) and also involved some parts of the inferior parietal lobule. In contrast, in recovery, which showed a stronger decrease than baseline, the most affected regions were the primary visual cortex in the occipital lobe and the superior parietal lobule, additionally, some small parts of frontal areas located anterior to the central fissure.

### 3.7.2.3. Temporal evolution of theta activity at sleep onset

Similar to alpha activity, mean theta activity (5-8 Hz) showed a saturating increase in the first 10 min after SO in both conditions reaching a plateau approximately at 5 min after SO. At the plateau, approximately 60% of the voxels differed between the two conditions (Fig. S3.7, left, green line) indicative of a faster buildup in recovery.

The percentage of voxels differing from the 2 min prior to SO was larger in recovery than in baseline. Between 70 to 75% of voxels differed in all intervals after SO (Fig. S3.7, right, solid lines), whereas in baseline a gradual buildup reaching 30-40% was present. Comparing consecutive 2-min intervals, only the first two intervals (1 and 3 min) differed from the previous one, approximately 20% of the voxels in baseline and 50 to 75% in recovery (Fig. S3.7, right, dashed lines).

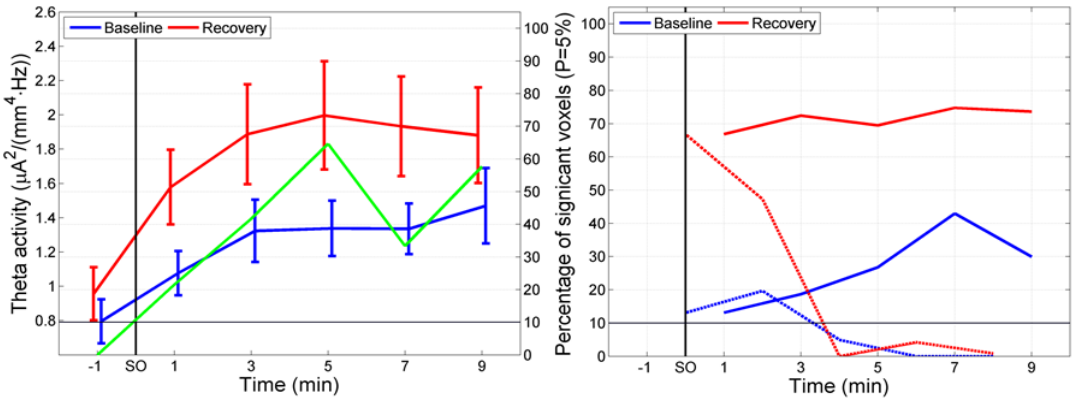


***Fig. S3.7.*** *Left panel: Temporal evolution of theta activity (mean squared magnitude of current source density in the theta band (5-8 Hz) [μA²/(mm⁴·Hz)]; averaged across all voxels; left y-axis) and statistical evaluation (percentage of significantly different voxels between conditions; right y-axis) at the transition to sleep (sleep onset, SO). Right panel: Percentage of voxels significantly different from the 2-min interval prior to SO (solid lines) and percentage of voxels significantly different between consecutive 2-min intervals (dashed lines).*

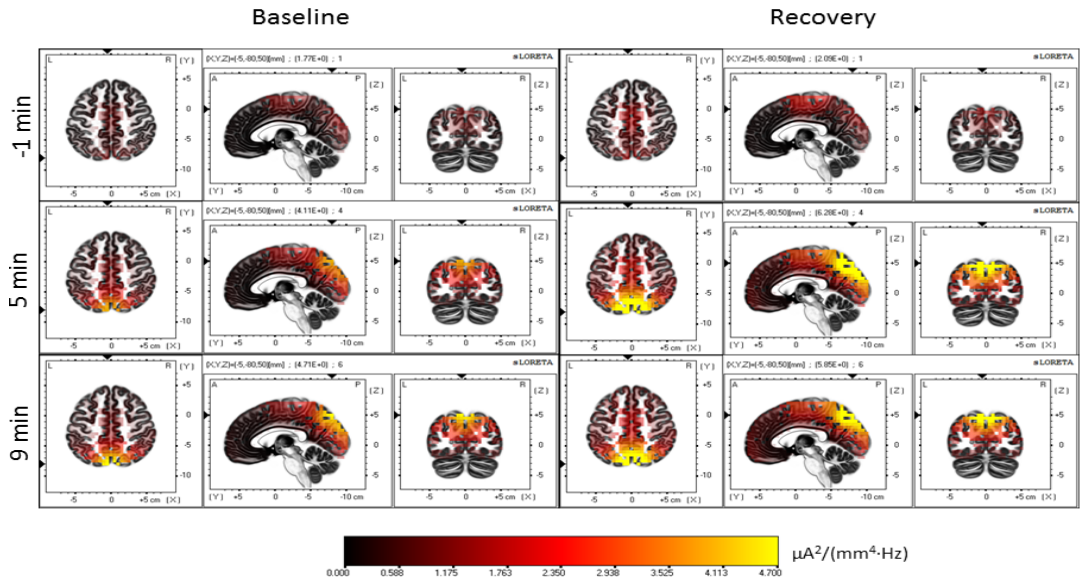


***Fig. S3.8.*** *LORETA source estimation in the theta band (squared magnitude of current source density [μA²/(mm⁴·Hz)] in the theta band (5-8 Hz)). Left: baseline, right: recovery sleep. Images represent average values of two minutes before SO (-1 min), 4-6 min (5 min) and 8-10 min (9 min) after SO. The plane coordinates were chosen to best match the most active brain regions (-5, -80, 50 in MNI coordinates [mm]).*

Fig. S3.8 depicts the gradual spatial buildup of theta activity (two minutes before SO (-1 min), 4-6 (5 min) and 8-10 min (9 min) after SO), both, in baseline (left) and recovery sleep (right). The neuroanatomical region with highest theta activity in both conditions was the precuneus (BA 7), but there were also contributions of the cuneus in the occipital lobe (BA 19) and, to a lesser degree, the postcentral gyrus in the parietal lobe (BA 3).

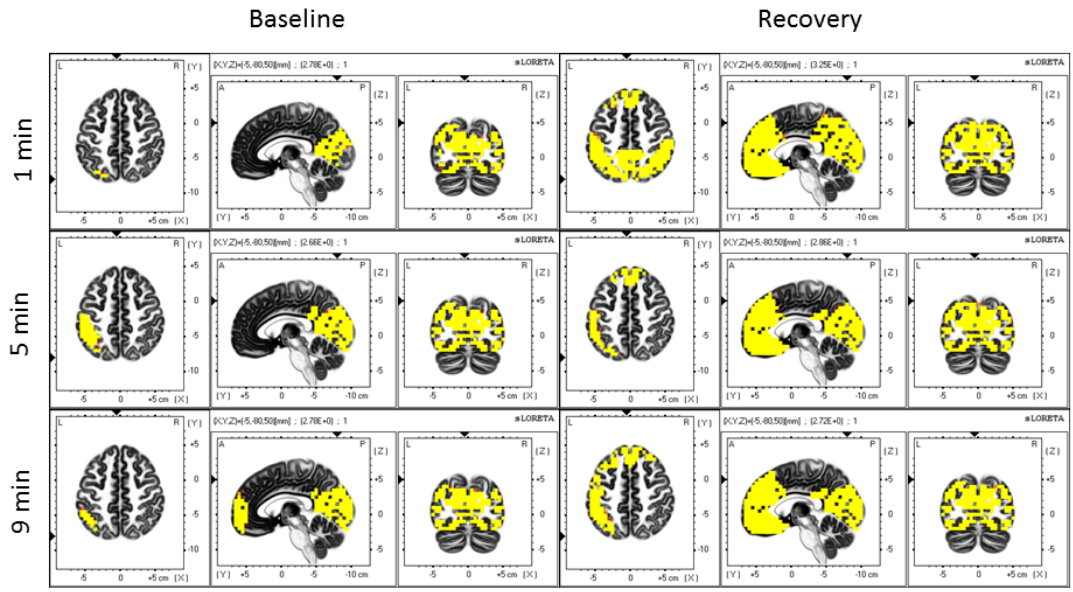


***Fig. S3.9.*** *Statistical evaluation of the temporal evolution of theta activity (5-8 Hz). Left: baseline, right: recovery sleep. Yellow areas represent voxels that showed higher theta activity after SO (t-values at P<0.05). Minutes 0-2 (1 min), 4-6 (5 min) and 8-10 (9 min) were compared to the two min prior to SO (-2 to 0 min). Coordinates of the planes are as in Fig. S3.8.*

Fig. S3.9 shows the statistical comparisons of minutes 0-2 (1 min), 4-6 (5 min) and 8-10 (9 min) with to the two min prior to SO (-2 to 0 min). The upper row (comparing 1st 2 min after SO with 2 min prior to SO) indicated for both conditions an increase in theta activity in large parts of the occipital lobe. However, recovery showed an even higher occipital activiation (basically covering the entire occipital cortex), and additionally, the participation of large areas in the prefrontal cortex (BA 9 and 10). After 5 min, larger occipital contributions were evident in baseline, whereas recovery remained rather constant. Finally, at 9 min, a moderate increase in the frontal lobe (orbitofrontal and medial prefrontal cortex) emerged in baseline, whereas recovery again remained basically constant.

# 4. Assessing intracortical causal information flow of oscillatory activity at the sleep onset transition with the isolated effective coherence (iCOH)

## Abstract

This research investigated the sleep onset transition in humans from an effective connectivity perspective. Using EEG recordings, source localization (LORETA) allowed us to reconstruct the underlying patterns of neuronal activity in various brain regions, e.g. the default mode network (DMN), dorsolateral prefrontal cortex and hippocampus, which were defined as regions of interest (ROI). We applied isolated effective coherence (iCOH) to assess effective connectivity patterns at the sleep onset transition. ICOH reveals directionality aspects and resolves the spectral characteristics of information flow in a given network of ROIs in an unprecedented fashion compared to other alternatives. We observed an anterior-posterior decoupling of the DMN, and moreover, a prominent role of the posterior cingulate cortex guiding the process of the sleep onset transition, particularly, by transmitting causal flow in the low frequency range (delta and theta bands) to other nodes of DMN (including the hippocampus). In addition, the midcingulate cortex appeared as a major cortical relay station for spindle synchronization (originating from the thalamus). The inclusion of hippocampus proved that this region was functionally involved in spindle synchronization observed in the cortex after sleep onset. Furthermore, under conditions of increased homeostatic pressure, we showed that an anterior-posterior decoupling of the DMN occurred, first, at a faster rate compared to baseline, and secondly, overall indicating weakened connectivity strength within the DMN. Finally, we also demonstrated that cortico-cortical spindle synchronization was less effective after sleep deprivation than in baseline, thus, reflecting the reduction of spindles under increased sleep pressure.

**Keywords:** effective connectivity, default mode network, delta activity, sigma activity, transition into sleep, sleep deprivation, isolated effective coherence, LORETA.

## 4.1. Introduction

Falling asleep is a process characterized by the relative disengagement from the external environment and a loss of consciousness, yet, little is known about its neurophysiological basis (Ogilvie, 2001, Marzano et al., 2013). Given that conscious awareness is attributed to the capability to integrate activity originating from diverse brain regions (particularly cortico-cortical and cortico-thalamic interactions), a suitable approach to examine how this process unfolds is by applying connectivity methods which are able to describe how distinct brain oscillations transmit information between brain regions (Massimini et al., 2005, Esser et al., 2009, Chow et al., 2013, Tononi et al., 2016). With a seed based approach or by independent component analyses (ICA), a number of brain networks were identified during rest (resting state networks, RSNs), with nodes clustered together according to a shared pattern of temporal correlations (Park and Friston, 2013, Amico et al., 2014). Two prominent networks stand out, the default mode network (DMN, thought to be involved in mind-wandering, creativity and emotional processing) and the central executive network (CEN, involved in cognition, planning and working memory), generally being anti-correlated to the DMN (Buckner et al., 2008, Douw et al., 2014, Salone et al., 2016). This study aims to investigate connectivity changes paralleling the sleep onset (SO) transition, on a selection of nodes that combines these two principal networks, which have been previously related with consciousness and sleep (Sämann et al., 2011, Heine et al., 2012, Uehara et al., 2014).

Identifying potential interactions arising in a network of connected brain regions is one of the most important and relevant problems in neuroscience (He and Evans, 2010, Park and Friston, 2013). In this context, connectivity can be addressed in three major ways by addressing structural, functional and effective connectivity (Sporns and Betzel, 2016, Stam et al., 2016). Briefly summarized, structural connectivity deals with the topology of white matter tracts physically linking different brain areas, functional connectivity computes temporal correlations between regions, and effective connectivity completes and corrects the deficiencies of functional connectivity, by

pruning potential indirect connections and by adding causal and directionality aspects (Friston, 2011, Valdes-Sosa et al., 2011, Stam et al., 2016). The use of the EEG, as in this study, and in contrast to fMRI research (typically limited to the mere predominant directionality of slow components due to the low temporal resolution of fMRI), allows for an additional spectral representation of effective connectivity coupling due to its much higher characteristic sampling rate (Pascual-Marqui et al., 2011). In this paper, we will focus on effective connectivity to study the SO transition from a connectivity perspective.

Effective connectivity methods are mathematical tools created to best characterize the influence that one neuronal group is impinging on another one in the context of neuronal networks (in principle, composed of an arbitrary number of nodes) (Friston, 2011, Liu and Aviyente, 2012). Pioneers in the study of causal relationships among a set of signals were Akaike (with a method called Noise Contribution Ratio) and Granger (with a causality method named after him), laying the analytical foundation for this field, which has advanced substantially since then, yet, continuing to be under development (Akaike, 1968, Granger, 1969). In general, effective connectivity methods can be formulated either in the time or frequency domain and can have either linear or non-linear characteristics in the underlying equations (Geweke, 1982, Bakhshayesh et al., 2014, Khadem and Hossein-Zadeh, 2014). Some well-known effective connectivity methods are: Granger Causality (GC), Partial Directed Coherence (PDC), Transfer Entropy (TE), Phase Slope Index (PSI) or Dynamic Causal Modelling (DCM) (Barnett et al., 2009, Silfverhuth et al., 2012, Ewald et al., 2013, Friston et al., 2014).

In order to assess effective connectivity changes and major properties characterizing the SO transition, we applied isolated effective coherence (iCOH), introduced by Pascual-Marqui (Pascual-Marqui et al., 2014a). Intuitively, this method provides a normalized measure (between 0 and 1) of the strength of coupling between a pair of nodes in the frequency domain, which is sensitive to the directionality of the interaction by index permutation (i.e., *A* directs *B*, versus the opposite, which can vary remarkably, a distinction neglected by functional connectivity methods) (Kralemann et al., 2014). By mathematical

construction, iCOH prunes away any indirect path linking two particular brain areas, which could distort the results in a major way (especially for complex networks). For example, if node *A* directs *B*, and *B* directs *C*, but there is no direct connection between *A* and *C* (existing only indirectly), then, iCOH results preserve the actual causal order, whereas other techniques (e.g., Granger Causality or Directed Transfer Function) would tend to incorrectly indicate a direct connection between *A* and *C*, biasing the causal structure of a network (Baccala and Sameshima, 2001, Faes et al., 2010). In addition, the frequency representation of iCOH allows to study the relevant information contained in its spectral characteristics and can be used to deduce which frequency bands (indicated by the position of either the absolute or local maxima) are preferentially involved in the information transfer between a set of brain regions (Pascual-Marqui et al., 2014a).

## 4.2. Methods

### 4.2.1. Data description and preprocessing

The EEG data analyzed in this study are from recordings performed on 8 healthy young males in a previous study (Finelli et al., 2000, Finelli et al., 2001a, Finelli et al., 2001b). The EEG signals were sampled at 128 Hz (band-pass filter: 0.16-30 Hz; for additional details see (Finelli et al., 2001b)). The sleep stages were scored for 20-s epochs (C3A2 derivation) according to Rechtschaffen & Kales (Rechtschaffen and Kales, 1968). Artifacts were identified as described in (Finelli et al., 2001b). Additionally, 4-s epochs containing artifacts that could not be removed were replaced by the preceding epoch free of artifacts (12-s epochs were analyzed). The total number of recording channels was 27, placed evenly on the scalp according to the international 10/20 system. Although a technical reference (placed 5% rostral to Cz) was originally used in the recordings, in the subsequent processing, all EEG data were re-referenced to the average reference. EEG recordings were performed during baseline sleep, and recovery sleep following 40 hours of sustained wakefulness. For details on participants (inclusion, exclusion criteria; protocol) and recordings see (Finelli et al., 2000,

Finelli et al., 2001b). Effective connectivity data were computed by means of iCOH (implemented in eLORETA) in the following frequency bands: delta (0.5-5 Hz), theta (5-8 Hz), alpha (8-12 Hz), sigma (12-16 Hz) and beta (16-24 Hz).

Even when the SO process is a gradual phenomenon that unfolds continuously as a function of time (Ogilvie, 2001, Prerau et al., 2014), a working definition must be provided in order to statistically compare and analyze two types of states, i.e., the period preceding and following SO. Thus, SO was operationally defined as the first occurrence of an epoch of stage 2 NREM sleep. EEG epochs used for the multivariate autoregressive (MVAR) model fitting, on which the iCOH method relies, had a length of 12 s. Outputs were combined by averaging to obtain a single effective connectivity response, representing a 2-min interval per subject, and subsequently averaged across subjects to yield a final group result. In order to have a consistent window length between conditions, given that sleep latency is shorter in recovery sleep (4.1±1.2 vs. 12.1±1.6 min; SEM; $P<0.01$; Wilcoxon signed-rank test), before SO there was only a single 2-min interval with a contribution of all participants. On the other hand, after SO, ten minutes were selected, resulting in five consecutive 2-min intervals, which were averaged to have a unified picture of the connectivity behavior both before and after SO. The eLORETA version used for analyses was version 20160611 (most updated at the time of running this study; available as free academic software from http://uzh.ch/keyinst/loreta).

### 4.2.2. Isolated effective coherence (iCOH)

The algorithm to compute iCOH starts with the estimation of a multivariate autoregressive (MVAR) model. For details on the mathematical deduction of the iCOH method based on a MVAR model, explained step by step, we refer to (Pascual-Marqui et al., 2014a). Since this effective connectivity analysis is performed at the source level (producing spurious results at the scalp level), it is necessary to use a source localization method for the computation of iCOH (Grech et al., 2008, Jatoi et al., 2014). The equation representing iCOH (derived in the frequency domain) takes the following expression:

$$\kappa_{i\leftarrow j}(\omega) = \frac{S_{\varepsilon_{ii}}^{-1}\left|\check{A}(\omega)_{ij}\right|^2}{S_{\varepsilon_{ii}}^{-1}\left|\check{A}(\omega)_{ij}\right|^2 + S_{\varepsilon_{jj}}^{-1}\left|\check{A}(\omega)_{jj}\right|^2}$$

satisfying the normalization condition: $0 \leq \kappa_{i\leftarrow j}(\omega) \leq 1$. Here, $\kappa_{i\leftarrow j}(\omega)$ represents the iCOH value at a given frequency $\omega$ between regions of interest (ROIs) $j$ and $i$, where the arrow indicates that $j$ influences $i$ (both indexes range from 1 to $N$, the number of ROIs). The matrix $\check{A}(\omega) = I - A(\omega)$, with $I$ being the unit matrix of order $N$, relates to the matrix $A(\omega)$ derived by least square fitting of the MVAR model of order $p$ (estimated by the Akaike information criterion). The matrix $S_\varepsilon$ represents the covariance of the residual errors of the MVAR model in the frequency domain and improves the result by adding a weight proportional to the accuracy of the accompanying parameters. The software LORETA is able to compute automatically all necessary parameters in this equation (including the optimal order $p$ of the MVAR model; $p$=23 in our case), rendering an iCOH spectrum as output, when providing a given set of EEG data as input.

The particular expression of the iCOH formula (specifically, as seen in the denominator), aims to convey that all connections beside the one of interest ($j$ to $i$) are mathematically "severed". For details about mathematical properties satisfied by iCOH, as well as a comparison with another effective connectivity techniques (the generalized partial

directed coherence (gPDC)), we refer to (Pascual-Marqui et al., 2014a, Pascual-Marqui et al., 2014b).

### 4.2.3. Selection of ROIs

In order to study statistical changes in effective connectivity accompanying the SO transition, a suitable selection of ROIs needs to be made, aiming to capture the most relevant aspects from a neurobiological perspective, thus, constituting a crucial step in the process of analysis (Sämann et al., 2011). The inverse solution provided by eLORETA allows for ROI definition either by providing the Montreal Neurological Institute coordinates of seed regions or by direct selection of Brodmann areas by the user. A total of nine ROIs were selected: medial prefrontal cortex (MPFC), mid-cingulate gyrus (MCC), posterior cingulate cortex (PCC), bilateral inferior parietal lobule (IPL), bilateral dorsolateral prefrontal cortex (DLPFC) and bilateral hippocampus (H). The rational for these particular ROIs comes from their close relationship with two principal brain networks: the DMN and the CEN (Sporns et al., 2007, Sämann et al., 2011). Thus, the first five ROIs define major hubs of the DMN (first two in the anterior, and last three in the posterior part), whereas the dorsolateral prefrontal cortex is the most important hub of the CEN (Andrews-Hanna et al., 2010, Larson-Prior et al., 2011). On the other hand, the hippocampus is sometimes included in the definition of DMN regions, whereas other authors prefer to consider the hippocampus separately as a limbic structure (therefore, excluded from DMN analysis), which nevertheless is acknowledged to interact with cortical hubs of the DMN (e.g., during autobiographical recall) (Buckner et al., 2008, Horovitz et al., 2009, Sämann et al., 2010). Either way, the main reason for adding the hippocampus to our ROI selection is due to its relationship with sleep spindles occurring at the SO transition, as revealed by intracranial recordings (Nir et al., 2011, Sarasso et al., 2014).

In practice, ROIs were defined by a sphere surrounding the centroid of each ROI. Nevertheless, an unavoidable tradeoff is present, as a small radius would not give the best precision (given that noise effects in the LORETA calculation of spectral current generators do not cancel out when the number of averaged vectors is small), whereas a large radius

would tend to produce wrong results for sensitive interactions. In addition, if the radius is too big, the averaged current vector can be even close to zero in all directions. Thus, as compromise, the radius for each ROI was defined as 15 mm (Siclari et al., 2014). The cartesian coordinates (expressed according to the Montreal Neurological Institute system, MNI) representing the centroids of our selected ROIs are listed in Table 4.1.

| **Name of Region of Interest (ROI)** | **Abbreviation** | **X (mm)** | **Y (mm)** | **Z (mm)** |
|---|---|---|---|---|
| Medial Prefrontal Cortex | MPFC | 0.0 | 45.0 | 13.3 |
| Mid-Cingulate Cortex | MCC | 0.0 | -17.5 | 37.5 |
| Posterior Cingulate Cortex | PCC | 0.0 | -52.5 | 27.5 |
| Left Inferior Parietal Lobule | LIPL | -45.0 | -45.0 | 52.5 |
| Right Inferior Parietal Lobule | RIPL | 45.0 | -45.0 | 52.5 |
| Left Dorsolateral Prefrontal Cortex | LDLPFC | -37.5 | 40.0 | 25.0 |
| Right Dorsolateral Prefrontal Cortex | RDLPFC | 37.5 | 40.0 | 25.0 |
| Left Hippocampus | LH | -21.6 | -28.3 | -10.0 |
| Right Hippocampus | RH | 21.6 | -28.3 | -10.0 |

***Table 4.1.*** *ROIs selected for the effective connectivity analysis (iCOH) at the sleep onset transition. Indicated are the centroid positions of the ROIs in MNI coordinates.*

### 4.2.4. Presentation of results and statistical analyses

Results are presented in two ways in this paper. Firstly, as iCOH values computed by eLORETA, representing connectivity strength (effect size) as a function of frequency for any pair of ROIs (plots are arranged as a square matrix of order 9; Fig. 4.1-4.3). Columns represent "senders", while rows are "receivers" of information flow. The correspondence between ROI and position in the square matrix, which is the same for rows and columns, follows the same order as they appear in Table 4.1. The frequency axis ranges from 0.5 Hz up to the Nyquist frequency (64 Hz), in 0.5 Hz steps. The y-axis represents the iCOH values. Dots below the curves indicate differences between conditions *A* and *B* (e.g., recovery versus baseline; grey dots *A*>*B*; black dots *A*<*B*). Only if clear differences between conditions were present did we interpret the data. Secondly, significant differences between conditions *A* and *B* are plotted in matrix fashion for specific frequency bands (Figs. 4.4 and 4.5). Red squares indicate *A*>*B*, blue ones *A*<*B*. All comparisons were corrected for multiple comparisons (see below).

We carried out statistical tests based on the method of non-parametric randomization of the maximum statistic, which has the advantage of correcting for multiple testing, and of not relying on the assumption of any specific or exact statistical distribution (e.g., Gaussian, Student, Fisher; statistical non-parametric mapping, SnPM) (Nichols and Holmes, 2002, Faes et al., 2012). With SnPM, surrogate permutations (5000 in our case) are created rendering a histogram which provides the statistical threshold (Pascual-Marqui et al., 2014a).

## 4.3. Results

### 4.3.1. iCOH spectra

#### 4.3.1.1. Baseline condition

Fig. 4.1 illustrates the iCOH spectra before and after SO in the baseline condition (performed on a 2-min window prior to SO (red curve) and averaged five 2-min windows after SO (blue curve); group averaged across subjects). Information flow was clearly asymmetrical with the MCC and PCC as major sources, i.e. drivers and the other regions as receivers, i.e. targets.

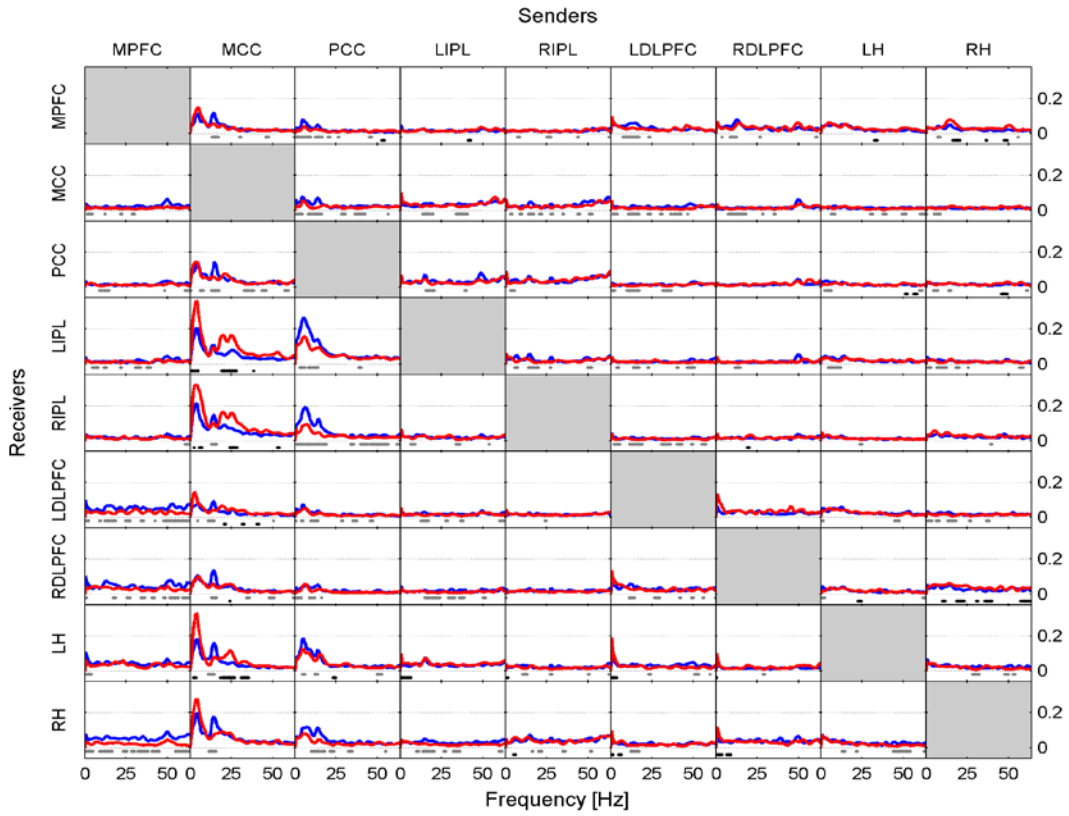


***Fig. 4.1.*** *Effective connectivity assessed by iCOH at the SO transition in baseline. iCOH values representing connectivity strength (effect size) as a function of frequency for any pair of ROIs. Plots arranged as a square matrix of order 9; columns represent "senders", rows are "receivers" of information flow. For the abbreviations of the ROIs see Table 4.1. iCOH before SO (red curves) and after SO (blue curves). Frequency range from 0.5 to 64 Hz (0.5-Hz resolution). Dots below the curves indicate significant differences between iCOH after SO vs iCOH before SO: grey dots iCOH after SO > iCOH before SO; black dots iCOH after SO < iCOH before SO.*

Before SO (Fig. 4.1, red curves), the MCC had the largest impact bilaterally on the inferior parietal lobule and hippocampus with the absolute maximum of iCOH located at 3.5 Hz (delta range) and to a smaller degree to the PCC and bilaterally to the dorsolateral prefrontal cortex. Information flow from the MCC to the MPFC occurred in the theta range (maximum at 5 Hz) and in the beta range (maxima at 20 Hz) bilaterally to the inferior parietal lobule and hippocampus. A further important source was the PCC, with predominantly theta (5.5 Hz) and to a smaller degree sigma (14.5 Hz) flow bilaterally to the inferior parietal lobule and hippocampus. These connections were less strong than those from the MCC.

After SO (Fig. 4.1, blue curves), one recognizes major changes that occurred after the transition into sleep compared to pre-SO. Now, the PCC achieved the greatest prominence as a driving source of information flow in the theta range, projecting bilaterally to the inferior parietal lobule (at 5.5 Hz (LIPL) and 6.5 Hz (RIPL)) and hippocampus (at 5.5 Hz (LH) and at 7.5 Hz (RH)). Further, the PCC projected to all other ROIs in the sigma range (related to spindles) with varying strength. The PCC is anatomically close to the thalamus, the actual generator of sleep spindles, thus, the PCC could be acting as one of its main cortical “relay stations” of spindle propagation (Amico et al., 2014). Secondly, after SO, the MCC was still exerting influence at various strength on all other ROIs mainly in the high delta (4 Hz), low theta (5 Hz) and in the sigma (14.5 Hz) range.

Dots in Fig. 4.1 represent the statistical comparison between the conditions, i.e., the changes that occurred after SO. Gray dots indicate frequency bins at which effective connectivity was enhanced after SO, black dots those at which it was decreased. The general picture reveals that prevalent connections present before SO originating from the MCC were weakened after SO in the low frequency range (delta and theta bands) and strengthened when originating from the PCC. In addition, causal flow in the beta range (located at 20 Hz and beyond) originating from the MCC was reduced. Finally, synchronization in the sigma band (sleep spindles) evolved mainly originating from the PCC and MCC.

### 4.3.1.2. Recovery condition

Fig. 4.2, similarly to Fig. 4.1, shows iCOH spectra before (red) and after (blue) SO in recovery after 40 h of sustained wakefulness. As in baseline, information flow was asymmetric with the MCC and PCC as major drivers. The connectivity patterns were like the ones observed in the baseline condition although with a different effect size (see below). Also, the comparison between post- and pre-SO revealed a similar picture of change as in baseline.

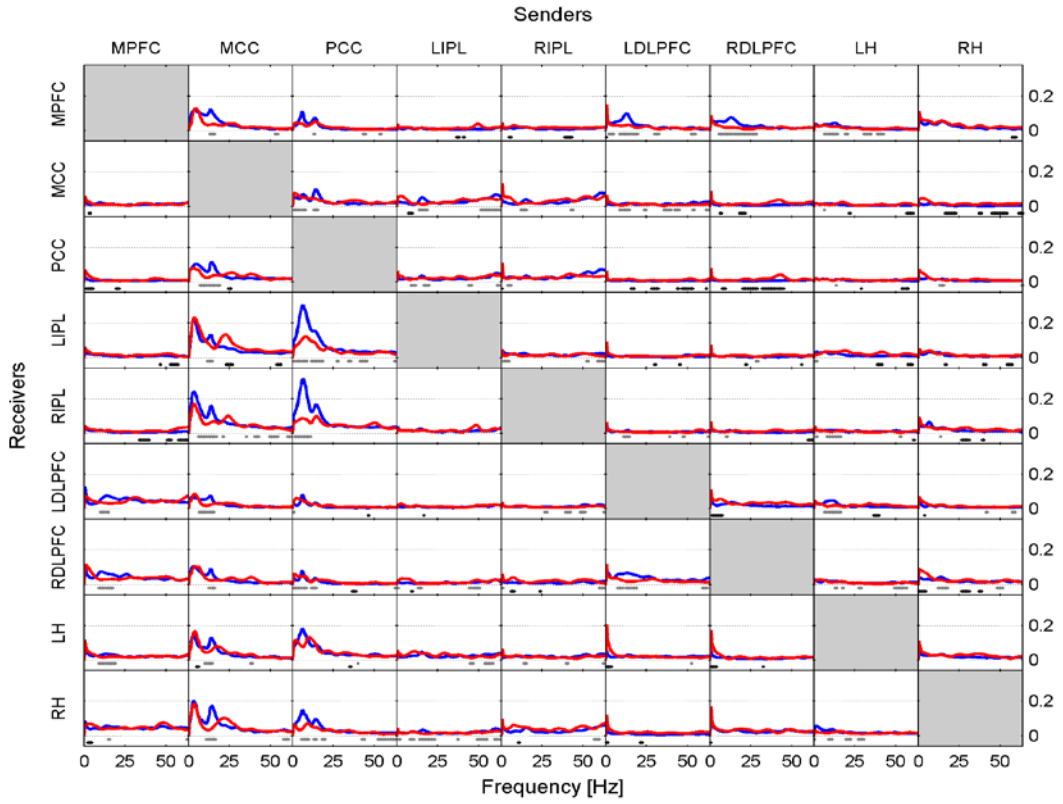


***Fig. 4.2.*** *Effective connectivity assessed by iCOH at the SO transition in recovery after 40 h of sustained wakefulness. For details of the representation see Fig. 4.1.*

#### 4.3.1.3. Statistical contrast between baseline and recovery conditions

Comparing iCOH spectra prior to SO between baseline and recovery revealed essentially no significant difference. Fig. 4.3 represents statistical comparison between recovery and baseline sleep for the period after SO, that is, comparing the blue iCOH spectra of Fig. 4.1 and 4.2.

The overall picture revealed a reduction of iCOH values (effect size) in recovery compared to the baseline following the SO transition (Fig. 4.3, black dots) across many pairs of ROIs. It affected broad frequency ranges (many of them with low effect sizes), often the beta and gamma range. Increased connectivity due to increased sleep pressure were observed for the following projections: mainly PCC to LIPL and RIPL (theta, alpha and sigma range), and MCC to MPFC, LIPL and RIPL (theta, upper alpha and sigma range). A few pairs of ROIs (out of a total of 72 connections) hardly showed any change: MPFC to RH, PCC to LH, RH to LH, and RIPL to MCC.

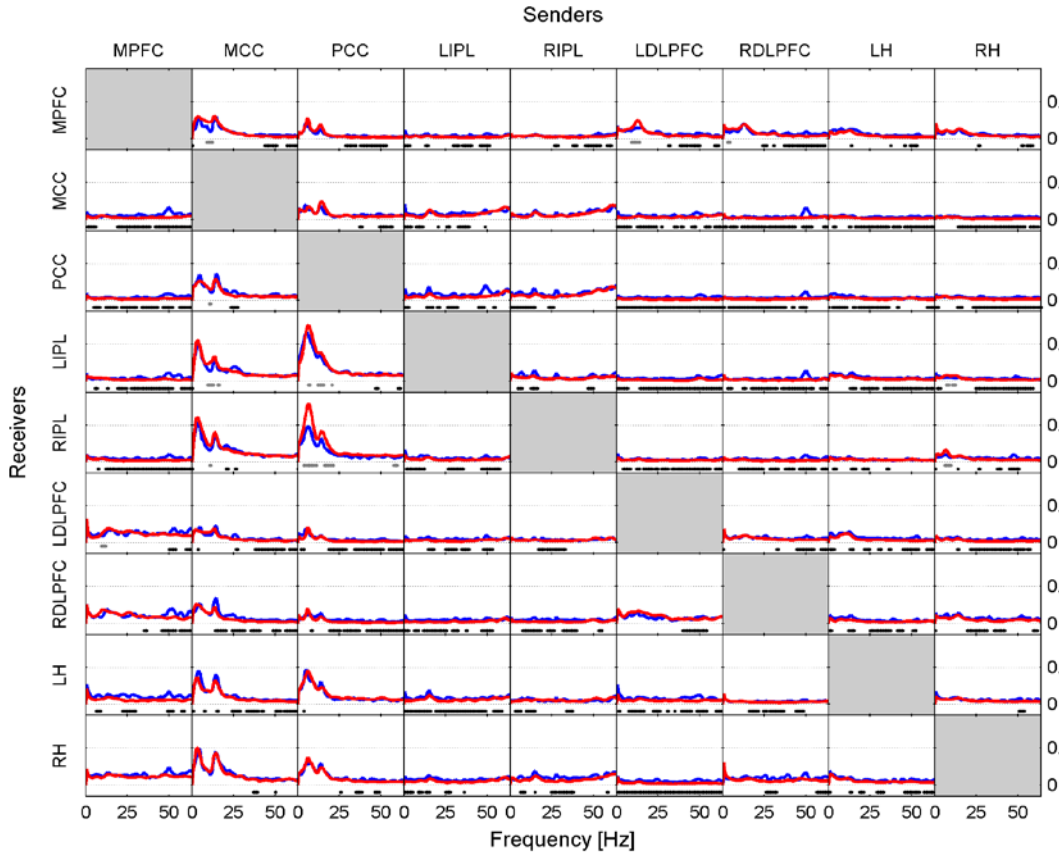


***Fig. 4.3.*** *Effective connectivity assessed by iCOH after the SO transition in baseline (blue curves) and recovery (red curves). For details of the representation see Fig. 4.1. Dots below the curves indicate significant differences between iCOH after SO in baseline vs iCOH after SO in recovery: grey dots iCOH recovery > iCOH baseline; black dots iCOH recovery < iCOH baseline.*

### 4.3.2. iCOH of specific frequency bands

We also investigated causal information flow in classical frequency bands. Fig. 4.4 illustrates significant effective connectivity changes following SO in the delta (0.5-5 Hz) and sigma (12-16 Hz) band (left and right columns, respectively) for baseline and recovery (upper and middle rows) and the contrast recovery to baseline (bottom row). Changes in effective connectivity of the theta (5-8 Hz), alpha (8-12 Hz) and beta (16-24 Hz) band are illustrated accordingly in Fig. 4.5.

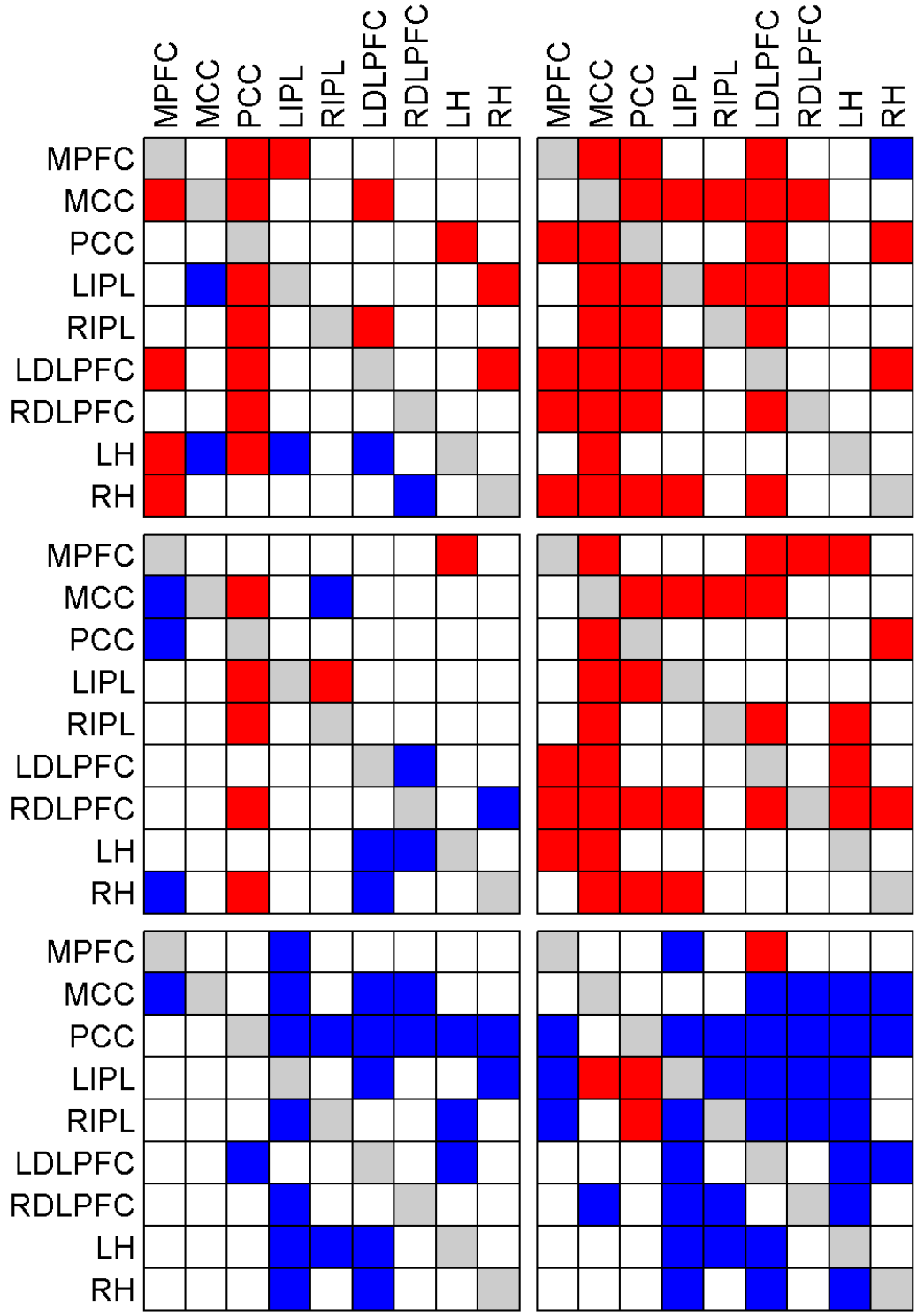


***Fig. 4.4.*** *Significant change of iCOH (connectivity) in the delta (left column) and sigma band (right column) at the SO transition in baseline (top) and recovery (middle) and after SO between baseline and recovery (bottom). Significant differences between conditions are plotted in matrix fashion; columns represent "senders", rows are "receivers" of information flow. For the abbreviations of the ROIs see Table 4.1. Red squares indicate connectivity before SO < connectivity after SO (top two rows) and connectivity at baseline < connectivity at recovery (bottom row); blue squares indicate the opposite change.*

#### 4.3.2.1. Delta band

In baseline, delta activity showed significant changes in information flow between 22 pairs of ROIs after SO, 17 increased and 5 decreased (Fig. 4.4, upper left panel). The PCC acted as a major low frequency synchronizer, connectivity to all ROIs except to the RH was increased (most significant increases occurred to the MPFC and MCC). On the other hand, the MCC, the other major source, showed weakened connections to the LIPL and the LH. All other changes concerned low effect sizes. Connectivity between the MPFC and the PCC, the LDLPFC, the LH and the RH, between the LIPL and the MPFC, between the LDLPFC and the MCC and the RIPL, between the LH and the PCC, and the RH and the LIPL and the LDLPFC was strengthened. Decreases (weakened connections) occurred from the MPFC to the LH, from the LDLPFC to the LH, and from the RDLPFC to the RH.

In recovery, information flow between only 16 pairs of ROIs changed (6 pairs less than in baseline), 7 increased and 9 decreased (Fig. 4.4, left middle panel) after SO in the delta band. The PCC continued to be a main driver of delta flow (however, displaying less significant connections), with increased connectivity strength to the MCC, bilaterally to the IPL, to the RDLPFC (highest statistical significance) and to the RH. Connections from the MCC to all other ROIs did not change. Weak connections changed as follows: The RIPL showed increased connectivity with the LIPL, and the LH with the MPFC. Decreased connectivity was observed between the MPFC and the MCC, the PCC and the RH, between the RIPL and the MCC, between the LDLPFC and the LH and RH, between the RDLPFC and the LDLPFC and the LH, and between the RH and the RDLPFC.

Comparing baseline and recovery, significant changes in connectivity basically only occurred after SO. The general picture revealed decreased effective connectivity in the delta band with increased sleep pressure (Fig. 4.4, lower left panel). Twenty-three pairs of ROIs showed decreased iCOH values during recovery compared to baseline. Thus, connectivity strength was higher in baseline. The MCC and PCC as major drivers did not show any change in connection strength with increased sleep pressure except for a reduction in information flow

from the PCC to the LDLPFC. All other reductions concerned iCOH values of low effect size.

#### 4.3.2.2. Sigma band

In baseline, 36 significant changes in connectivity, 35 of them representing an increase in the information flow with SO in the sigma band (Fig. 4.4, upper right panel). The MCC occurred to be the main cortical driver of sigma flow in our selection of ROIs, with significantly increased projections to all other ROIs and overall displaying the highest significant increases with the SO transition. Similarly, the PCC also appeared as an important sigma driver (although less than the MCC), directing connections to all ROIs, except to the LH. All other connections showed weak (low effect size) to very weak changes. The MPFC increased transmission to the PCC, bilaterally to the DLPFC and to the RH; the LIPL to the MCC, the LDLPFC and the RH; the RIPL to the MCC and the LIPL; the LDLPFC mainly to its right counterpart and the MPFC, and weakly to the MCC and PCC, bilaterally to the IPL and the RH; the RDLPFC mainly to the MCC. Finally, the RH decreased sigma transmission to the MPFC, but increased it to the PCC and LDLPFC.

In recovery, 30 connections showed increased information flow between pairs of ROIs (6 less than in baseline) after SO (Fig. 4.4, middle right panel). Most of them overlapped with those of baseline some disappeared and a few weak ones that occurred with higher sleep pressure. Again, with the MCC and PCC being the main cortical drivers of sigma flow.

Connectivity in baseline and recovery basically differed only after SO. Generally, decreased effective connectivity with increased sleep pressure was observed between 35 pairs of ROIs and increased connectivity between 4 pairs (Fig. 4.4, lower right panel). The MCC exerted a stronger impact on the LIPL, and a reduced one on the RDLPFC. The PCC showed heightened information flow bilaterally to the IPL. Most other changes concerned weak connectivity.

#### 4.3.2.3. Additional frequency bands

Changes in effective connectivity of the theta, alpha and beta band are illustrated in Fig. 4.5. As a general trend, both the theta and the alpha band undergo a significant increase of effective connectivity with SO. For the theta band, major drivers are the PCC and the hippocampus (bilaterally). In the case of the alpha band, the major drivers comprise the PCC and all defined ROIs located on prefrontal regions (MPFC and bilateral DLPFC). Additionally, the alpha band was more resilient to the effects of sleep deprivation on effective connectivity, as it still shows a considerable number of statistical changes, closer to the baseline condition. On the other hand, results for the beta band are of less interest, as only a small number of connecting pairs show significant changes with SO. Nonetheless, we also observed in the beta band a significant reduction of effective connectivity in the recovery case, thus, similar to the other bands in this regard.

In the theta and alpha bands at baseline (top row), mostly increased effective connectivity was observed. In the alpha band, the PCC was a major driver of causal flow at the transition. These changes were less obvious in the beta band (fewer connectivity changes with SO), although they generally abide a similar dynamic. In recovery (middle row), qualitatively similar changes as in baseline occurred in all bands, although with fewer changes than in baseline. Nonetheless, the PCC could still be identified as the most important ROI causally affecting other ROIs. Comparing baseline and recovery (bottom row), an overall reduction of iCOH occurred after SO in all frequency bands (in the alpha band, there were several exceptions to this general tendency). We may interpret the last contrast as an indication of a reduced need of connectivity reorganization in a sleep deprived brain while falling asleep due to an increased homeostatic drive (congruent with a larger number of significant changes at baseline than in recovery). For more information on Fig. 4.5, we refer to the Supplementary Material (see section 4.5.4).

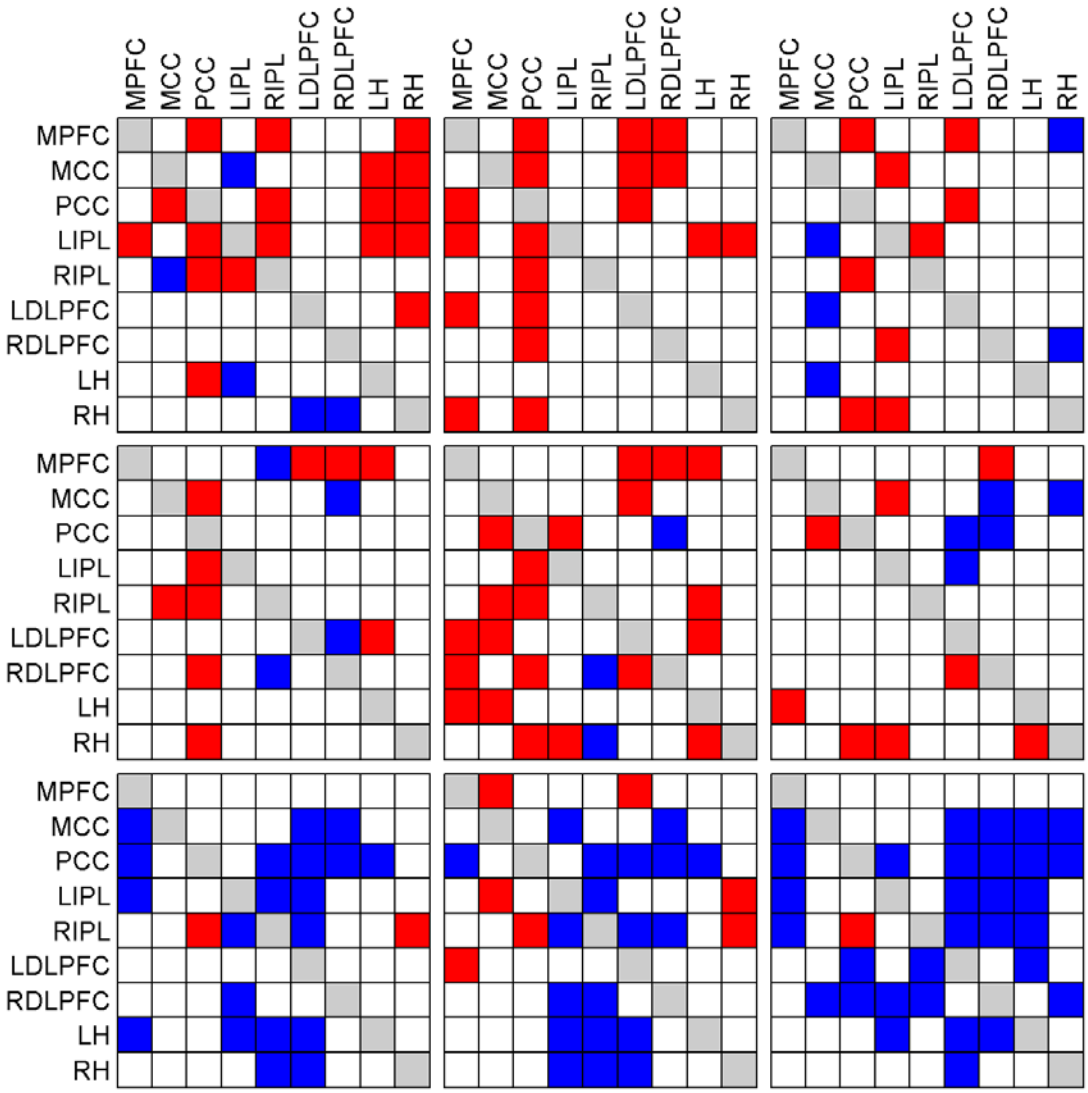


***Fig. 4.5.*** *Significant change of iCOH (connectivity) in the theta (left column), alpha (middle column) and beta band (right column) at the SO transition in baseline (top) and recovery (middle) and after SO between baseline and recovery (bottom). For details see Fig. 4.4.*

## 4.4. Discussion

Overall, we found significant effective connectivity changes, as assessed by iCOH, during the SO transition, which are congruent with the role of the PCC (a major hub of the DMN) in regulating consciousness (Amico et al., 2014, Herbet et al., 2014). In addition, we showed that, in the cortex, the MCC (and to a lower extent, the PCC) acted as the main relay stations of the thalamus for fast spindle synchronization after SO (Andrillon et al., 2011).

### 4.4.1. Relevant spectral features of iCOH

#### 4.4.1.1. Baseline

Low-frequency transmission arising from the MCC, and to a lesser extent from the PCC to the hippocampus and posterior parts of the DMN appear to be the major causal drivers characterizing the period prior to SO in the baseline condition (Fig.4.1, red curves). After the SO transition, the PCC emerged as the most salient structure and principal driver of synchronization, sending theta flow to basically all investigated ROIs (Fig. 4.1, blue curves and Fig. 4.4, top row). The MCC still exerted a prominent role (although, in relative terms, to a smaller extent than prior to SO), transmitting causal flow at slightly higher frequencies (high delta/low theta) than before SO. After SO, spindle synchronization (sigma band) occurred throughout all defined ROIs, driven by both the MCC and the PCC. However, that does not necessarily imply spindles must be described as a global cortical phenomenon, as local contributions from fronto-parietal areas might be the most important (Andrillon et al., 2011, Del Felice et al., 2014).

The DLPFC bilaterally also exhibited transmission flow, particularly to the hippocampal complex displaying absolute maxima at very low frequencies. Nevertheless, this effect is most likely due to window effects happening near 0 Hz (attributed to the Gibbs phenomenon in signal processing theory), smearing a bit for low delta frequencies (as the iCOH plots are smooth, coming out of a MVAR model), rather than being a neurobiological effect (Wanguemert-Perez et al., 2007). The window effect behavior was also observed for transmission flow out of the RH (in recovery).

#### 4.4.1.2. Recovery

Prior to SO, a delta and theta flow from the MCC (acting as principal source of synchronization), directed to the remaining ROIs was observed (Fig. 4.2, red curves). There was an additional beta flow component from the MCC PCC, mainly projecting bilaterally to the inferior parietal lobule and hippocampus.

After SO, firstly theta flow originating from the PCC as main driver, and a secondary delta flow arising from the MCC was present (Fig. 4.2, blue curves). Secondly, spindle synchronization spread across all ROIs (sinks), originating from the MCC and PCC, hence, localizing cortical sources of fast spindles compatible with previous studies (Del Felice et al., 2014, Park et al., 2015). Furthermore, spindle synchronization displayed higher clustering in the frontal lobe during recovery than in baseline, as shown by the enhanced dorsolateral prefrontal contribution (compare Fig. 4.2 with Fig. 4.1; see also Fig. 4.4 for more information). Additionally, a pair of connections not present in baseline (Fig. 4.1) from the bilateral DLPFC (bilaterally) to the MPFC, peaking at 13 Hz, thus, again reflecting local spindle synchronization throughout the prefrontal cortex. In particular, the dorsolateral prefrontal cortex (a cortical structure well-known to produce enhanced delta activity during recovery sleep) also received low frequency input, thereby, possibly facilitating the generation of slow waves (De Gennaro et al., 2000a, Marzano et al., 2013). It is also worth to notice the sharp disappearance of the beta flow (Fig. 4.2) in the period after SO, which was present before SO, an effect that was less pronounced in baseline (Fig. 4.1).

In general, the changes at SO in recovery were qualitatively similar to baseline (main differences pertain to the iCOH levels, i.e. effect size), confirming the relevance of the MCC as predominant source of spindle synchronization, with a secondary contribution of the PCC. Local spindle connectivity along the prefrontal cortex as revealed by bidirectional connections originating from the MPFC and dorsolateral prefrontal cortices, thus, constituting a network cluster in the frontal lobe, but also communicating with centro-posterior structures (Fig. 4.4). The hippocampus, particularly the left part, showed a not

negligible role at the SO transition, with transmission flow in a broad frequency range to the frontal and parietal lobes.

#### 4.4.1.3. Recovery versus baseline

Most evident was a reduction in connectivity after SO in recovery compared to baseline (Fig. 4.3, black dots). This effect was most noticeable in the beta range (starting at 16 Hz, if the gamma range is also subsided as high beta), conveying the notion of significantly less beta flow in the sleep deprived brain (associated with cortical and body arousal, thus, fitting with decreased arousal in this state), although not exclusively restricted to the high frequency range (Kuo et al., 2016, Sena et al., 2016). We can reasonably attribute the significant reduction of strength in cortical communication to the breakdown of effective connectivity accompanying SO (Massimini et al., 2005, Esser et al., 2009), which based on our results, would be even more prominent in recovery sleep. Nevertheless, there were also connections undergoing a local increase in recovery at certain frequencies. In general, the frequency bands with increased connectivity pertained to either the upper alpha or low sigma range, potentially reflecting spindle synchronization, although we cannot rule out smearing from the upper alpha range.

### 4.4.2. Topographical properties and neurobiological interpretation

Changes in effective connectivity in the delta band paralleling the SO transition indicated a greater cortical breakdown affecting fronto-parietal DMN nodes during the recovery condition (Fig. 4.4, left column) (Massimini et al., 2005, Bonhomme et al., 2016). Indeed, integrity of the DMN, particularly the coupling between anterior and posterior parts, has been associated to the degree of conscious arousal or alertness (e.g., it is strengthened in practice of mindfulness aimed to boost attention and weakened in ADHD) (Horovitz et al., 2009, Chow et al., 2013, Franzen et al., 2013, Britton et al., 2014). For instance, functional connectivity in fMRI studies (based on linear temporal correlations analyses, using DMN nodes as seed regions) applied to the SO transition have also indicated decoupling of anterior and posterior parts of the DMN with increased sleep depth (Sämann et al., 2011).

Consequently, the gradual disruption of anterior to posterior parts of the DMN shown by iCOH might correlate at a subjective level with the progressive fading of consciousness during SO (Heine et al., 2012, Speth and Speth, 2016).

Our effective connectivity analyses also indicated that this disruptive process occurs at an accelerated rate when the brain is exposed to increased sleep pressure, a behavior also observed with functional connectivity in fMRI (based on cross-correlations, a technique far less rigorous than ours) (Sämann et al., 2010). Furthermore, the role of the PCC for regulating consciousness and controlling DMN rearrangement in the context of SO has also been suggested using fMRI, although it could not be rigorously proven simply based on seed correlations (Sämann et al., 2011). Thus, the higher rate of DMN decoupling occurring under sleep deprivation (Fig. 4.4, left column) might occur in baseline later in time during slow wave sleep. Additionally, disruption of the MPFC to PCC coupling has also been reported in vegetative state patients and during deep anesthesia. Thus, Fig. 4.4 would indicate that the reversible “impairment” of consciousness defining the sleep state is more profound during the recovery condition, as a result of reaching a higher network disconnection than in baseline (Boly et al., 2008, Barrett et al., 2012). Hence, the fronto-parietal disconnection (Fig. 4.4; left middle panel) may be seen as a temporally accelerated connectivity change reaching a deeper loss of conscious awareness.

Effective connectivity analyses pointed to the PCC as the major hub affecting effective connectivity breakdown during SO. Indeed, this structure has been shown in many studies to play a crucial, necessary, albeit possibly not sufficient, role in maintenance of a normal state of consciousness, or alternation between different states of consciousness (including pathology) (Vogt and Laureys, 2005, Amico et al., 2014, Herbet et al., 2014, Leech and Sharp, 2014). Therefore, we consider that this finding fits nicely and elegantly with the quintessential process of consciousness fading during SO. For instance, experimentally induced disruption of the PCC in epileptic patients generates subjective feelings of depersonalization (i.e., decreased agentive self-awareness, or “feeling like a doll”) and a loss of external connectedness (e.g., feeling of an oneiric world), suggesting that

functional integrity of this structure is necessary for sustained consciousness related to the external environment and, perhaps, body-schema (Herbet et al., 2014, Herbet et al., 2016). Moreover, the PCC plays a crucial role in the normal DMN functioning, and its relevance is highlighted by the fact that it has the highest metabolic rate of the brain and the highest degree of functional connectivity to other brain areas in resting-state analyses performed with fMRI studies (Vogt and Laureys, 2005, Sämann et al., 2011).

Hallucinations, either visual, auditory or even somatosensory, reported after electrical stimulation of the PCC can be compared to hypnagogic hallucinations occasionally occurring during SO, hence, we speculate that there might be a partially common neurobiological basis for both phenomena (Magnin et al., 2010). In fact, the PCC acts as an integrative hub of information convergence from the hippocampus and MPFC (Leech and Sharp, 2014). As a result, it is likely that these hallucinations are "inspired" by biographical memories collected throughout the day in the hippocampus; moreover, the MPFC is congruent with selection of self-referential meaning (Buckner et al., 2008, Esslen et al., 2008a). Hypnagogic hallucinations might more likely occur during baseline than recovery (due to its higher anterior to posterior disconnection and weakened involvement of the hippocampus, Fig. 4.4, upper and middle left panels), in agreement with subjective experience.

Creative states known to often occur during the SO transition, which have been credited throughout history by many scientists, artists and other creative achievers (indeed, the frequency of this phenomenon is greater among creative individuals) (Jimenovaldes and Jimenobulnes, 1993, Watanabe, 1996), could also be partially explained by our effective connectivity results. In this way, transmission at low frequencies (mostly in the theta range) from the PCC (a key region for fluency of idea generation, positively correlated with amount of gray matter in this region) and the disengagement of the MPFC (an area that reduces BOLD activity during musical improvisation and flow states) are common to SO and flow states (Kuhn et al., 2014, Ulrich et al., 2014, Jauk et al., 2015). A possible interpretation to these shared

neurobiological aspects could be the sense of “vanishing self” that is subjectively described both during flow states and SO, as self-awareness might be critically supported by the major hubs of the DMN (i.e., MPFC and PCC) (Brewer et al., 2013, Ulrich et al., 2014). Thus, divergent thinking accompanying the SO transition through hypnagogic hallucinations and external disengagement, towards a more inwardly oriented and disinhibited state (given the reduced integration of the prefrontal cortex with self-referential posterior structures, particularly, the PCC), could be matched to DMN connectivity reorganization during SO (Fig. 4.4) (Larson-Prior et al., 2011, Brewer et al., 2013). Nevertheless, this process is less likely to occur during recovery sleep given the much accelerated network disruption with an increased homeostatic debt (Fig. 4.4 middle row).

Finally, as another prominent feature the hippocampus received less delta flow from the prefrontal cortex in the recovery condition, and in turn, sent less flow itself (Fig. 4.4). The hippocampus is a structure known to be involved in both short-term and long-term memory, and has also been implicated in working memory in conjunction with the prefrontal cortex (Lavenex and Amaral, 2000, Leszczynski, 2011). Furthermore, the hippocampus is documented to play a pivotal role in transferring of declarative memories gathered during the day to the cortex for long-term storage, a process that happens principally during stage 2 and deep sleep (Born et al., 2006, Rasch and Born, 2013). We hypothesize that this prefrontal-hippocampal effective connectivity breakdown may constitute a neurobiological explanation for impaired declarative memory consolidation under conditions of sleep deprivation (Lavenex and Amaral, 2000, Gais and Born, 2004).

In general, the topology of the network indicating major statistical changes unfolding with the SO transition in the sigma range (Fig. 4.4, right column) exhibited a great similarity between the two conditions, although the number of statistically significant connections was a bit higher at baseline (36 significant connections, compared to 30 in the recovery condition). The lower number of significant connections in the recovery condition could indicate a reduced capacity to generate and propagate spindles throughout the cortical mantle when subjects are under higher sleep pressure (Olbrich et al., 2014). This may also

relate to the inverse relationship between delta and sigma activity (delta activity increasing with sleep pressure), and as such leading to less spindles in recovery sleep (Carrier et al., 2001, Finelli et al., 2001b). The topology was mainly governed by fronto-parietal connections, generally indicating a statistical increase in iCOH values, hence, in agreement with the occurrence of spindle synchronization with SO (De Gennaro et al., 2000a, Anderer et al., 2001). Thus, this suggests that spindles are not isolated, without a possibility to communicate, but rather, they spread across the cortical surface through cortico-cortical interactions, collectively aiding to induce synchronization in other cortical areas (Andrillon et al., 2011).

As we defined the sigma band as 12-16 Hz, effective connectivity results relate to fast (rather than slow) spindles (Aeschbach and Borbély, 1993, Aeschbach et al., 1997). Although spindles have a thalamic origin (LORETA cannot localize deep subcortical structures), as a relay, the major cortical hub acting as a driver for spindle synchronization (in this case, fast spindles) was the MCC, both in baseline and recovery conditions (in baseline, the PCC also exerted an important, but secondary, contribution) (Amico et al., 2014). Given that fast spindles show maximum power spectral density along the parietal lobe, the spatial location of MCC is optimal for being the driver of fast spindles (Anderer et al., 2001, Marzano et al., 2013). This structure is compatible with other studies of source localization using LORETA, but which were based solely on power analyses (Anderer et al., 2001, Andrillon et al., 2011, Del Felice et al., 2014). In the previous chapter, we also found through source localization significantly high sigma (spindle) activity after SO on a hot zone comprising posteromedial regions, which include and surround the middle and posterior cingulate gyrus (see Figs. 3.5 and 3.6). We think that an approach based on effective connectivity is methodologically superior than power spectral density for localizing the source of spindles, because the interactive aspects are totally neglected in power maps, an spindle emergence is a phenomenon in which synchronous coupling is essential (Saletin et al., 2013). Additionally, the fact that during baseline there were two main active sources (i.e., the MCC and PCC), whereas during recovery only the most important was active, is

congruent with observations that after SO, the cortical ability to induce and propagate spindles is hindered (De Gennaro et al., 2000a, Olbrich et al., 2014).

For both delta and sigma activity, the general trend, with few exceptions, was a reduction of connectivity strength in the recovery condition. The spatial organization of the network indicating significant changes remains remarkably similar in the two conditions, suggesting that the recovery condition is not characterized by a new spatial configuration, different from baseline, but rather by a loss of connectivity strength within the same main fronto-parietal networks. Significantly lower iCOH values in the delta band during sleep deprivation would correspond to a higher level of cortical breakdown of effective connectivity, and consequently, even more reduced propagation of waves throughout the cortex triggered by transcranial magnetic stimulation (Massimini et al., 2005, Esser et al., 2009). The PCC appeared again as the most densely interconnected ROI manifesting major statistical changes, suggesting that during SO it is "switching off" the neuronal underpinnings of consciousness that it previously helped to maintain, and doing so faster during recovery, given the reduced effective connectivity, or higher DMN disruption (Amico et al., 2014). We can expect that diminished connectivity strength observed in the recovery condition would occur later in the baseline condition as sleep continues to deepen (Sämann et al., 2010, Larson-Prior et al., 2011). Consequently, as it is also manifested by the faster build-up of delta activity during recovery sleep, the connectivity approach is also compatible with the interpretation of this state as a fast-forward version of the dynamical evolution in baseline, achieving greater decoupling between anterior and posterior parts of the DMN at a higher rate (Horovitz et al., 2009, Marzano et al., 2013).

In turn, significantly lower iCOH values in the sigma band, topologically ascribed to fronto-parietal networks (Fig. 4.4, lower row, where the cingulum plays a pivotal role), indicate impaired capacity to generate and sustain spindle activity with increased sleep pressure, in agreement with diminished sigma activity during recovery sleep, and possibly, with an interplay relationship between connectivity and activity (De Gennaro et al., 2000a). Cortico-cortical spindle

synchronous behavior is akin to synchronization of many metronomes, where the coupling allowing transmission is the supportive surface, and the more oscillators are already synchronous, the greater the induction of synchronization on other oscillators, until achieving a tight and common synchronous state. In this analogy, the recovery condition would erode the substrate for synchronization. Moreover, fewer and weakened spindle synchronization (i.e., less connections and lower iCOH values; Fig. 4.4, right column) during the recovery sleep could also explain the impaired consolidation of declarative memory in this condition, a process where spindles play a crucial role (Gais and Born, 2004, Born et al., 2006). For example, impaired declarative memory consolidation with aging was attributed to a lower spindle density, with subjects exhibiting better performance also displaying greater spindle density and, furthermore, declarative memory consolidation was boosted using transcranial direct current stimulation to increase spindle density (Seeck-Hirschner et al., 2012, Ladenbauer et al., 2016). Accordingly, the hippocampus exhibits a sharp disconnection from cortical areas (particularly the frontal lobe) in the recovery condition, which would hinder transferring of autobiographical and declarative information to the cortex for long-term consolidation (Nir et al., 2011, Mednick et al., 2013). Finally, the inclusion of the hippocampus in the iCOH analyses, which was shown to be an active player during SO (at least, during baseline), relates to the experimentally observed behavior of hippocampal spindles preceding neocortical spindles, as inferred from intracranial recordings in surgical patients (Sarasso et al., 2012, Sarasso et al., 2014), a claim that can be supported from our non-invasive connectivity perspective.

The anterior-posterior decoupling following the SO transition has also been observed using yet another effective connectivity technique different from the iCOH, the Direct Transfer Function (DTF) (De Gennaro et al., 2004). In this regard, the DTF showed, for the period preceding SO (defined as emergence of first spindle or K complex), a prevalence of occipital to frontal information flow in the delta, theta and alpha bands. However, after SO, the directionality pattern inverted, and the predominant direction of transmission was fronto-parietal to occipital at all frequency bands. Although this analyses was based on

the scalp EEG and did not define brain networks, the observed behavior is compatible with a breakdown of the default mode network (whose principal nodes are the medial prefrontal cortex, in the anterior part, and the posterior cingulate cortex, in the posterior part), particularly in the low-frequency range.

In addition, a later study confirmed and extended the previous finding based on the DTF (De Gennaro et al., 2005b). In this study, the effects on connectivity resulting from total sleep deprivation were also assessed by means of the DTF. With increased sleep pressure, the anterior-to-posterior directionality of coupling could be already detected before the SO transition, therefore, constituting a time advance shift compared to the baseline condition. This speed-up of the dynamics has also been observed in our study. In addition, the connection between electrodes Fz-Pz, showed significantly lower DTF values in the low frequency range (and up to 20 Hz) during recovery.

Finally, an fRMI study of functional connectivity and graph theory also concluded that the SO transition is accompanied by a breakdown of cortico-cortical connectivity as sleep progresses into slow wave sleep (Spoormaker et al., 2010). The breakdown in cortico-cortical connectivity between anterior and posterior nodes of the DMN serves as a neurophysiological explanation for disengagement from the external world, as it hinders the capacity of cortical areas to integrate information received from other brain areas. However, before entering into slow wave sleep, cortico-cortical connectivity was observed to be higher than in wakefulness, which fits with our data (in particular, synchronization of spindles, although the fMRI study lacks the tools to tackle such questions given the slow dynamics of the BOLD signal). Thus, connectivity was maximal during light sleep and increased with the SO transition, and subsequently diminished entering slow wave sleep. In addition, posterior hubs were observed to maintain or even strengthen their connectivity as sleep deepened, which suits with the predominant role of the posterior cingulate cortex observed in our analysis.

## 4.5. Supplementary Material

### 4.5.1. Brief explanation of iCOH formula and characteristics

The particular expression of the iCOH formula (specifically, as seen in the denominator), aims to convey that all connections beside the one of interest (*j* to *i*) are mathematically "severed". As a consequence, iCOH exhibits three major advantages. Firstly, iCOH is robust to the number of ROIs employed, whereas other normalizations would reduce (or rise) the values for increasing (or decreasing) number of ROIs investigated, even when the relationship between sender and receiver remains unaltered (e.g., this drawback appears in both the Partial Directed Coherence and Direct Transfer Function) (Blinowska, 2011). Secondly, iCOH discounts all potential indirect connectivity paths, i.e., results yielded by this method will only be attributable to truly direct connections (a major drawback of other methods, such as the Direct Transfer Function, Granger Causality or Akaike's Noise Contribution Ratio, which tend to mix direct and indirect connections altogether) (Philiastides and Sajda, 2006, Pascual-Marqui et al., 2011). Thirdly, iCOH provides a faithful frequency response, in the sense that, if normalized in a different way (such as the Partial Directed Coherence), the frequency response would be contaminated by information pertaining to other coefficients that are not relevant for the given pair of ROIs of interest (Pascual-Marqui et al., 2014a, Pascual-Marqui et al., 2014b).

As it is known, besides iCOH, there are several other effective connectivity techniques that can be used to address the same problem of coupling strength and directionality (Baccala and Sameshima, 2001, Blinowska, 2011, Silfverhuth et al., 2012). Therefore, the particular selection of iCOH, instead of other methods, can be justified based on the caveats found in these alternative measures (Sommerlade et al., 2011). Perhaps the most prominent linear alternative to iCOH is the above mentioned Partial Directed Coherence (PDC) (this method also corrects for indirect connectivity paths, thus, being frequently used), as well as its improved version (which adds variance errors as weighting coefficient factors for better accuracy), the generalized Partial Directed Coherence (gPDC) (Baccala and Sameshima, 1999, Astolfi et al.,

2007). Other popular alternatives known in the effective connectivity literature, such as Granger Causality (GC) and the Direct Transfer Function (DTF), can be discarded on the ground that PDC outperforms them by discounting indirect paths between nodes and only retaining direct paths, a major pitfall implicitly present in the mathematical formulation of these methods (related to the matrix inversion that yields the transfer function matrix) (Granger, 1969, Schelter et al., 2006, Blinowska, 2011). Another relevant alternative method, the directed Direct Transfer Function (dDTF) introduced by Korzeniewska et al., also corrects at the theoretical level, as iCOH, the issue of eliminating unwanted indirect causal paths, but in practice, it has been shown to systematically render excessive low values (showing low sensitivity, and in many cases, being statistically indistinguishable from noise), thus, justifying its exclusion (Korzeniewska et al., 2003, Astolfi et al., 2007). Consequently, it will be sufficient to explain why the iCOH outperforms the gPDC.

To this aim, a computational toy example was proposed by Pascual-Marqui et al. to compare the results provided by each of these methods (Pascual-Marqui et al., 2014b). In this way, several signals (covering five different nodes arranged in a designed network diagram) were artificially simulated as if they were extracted from real EEG data, hence, having the advantage that the actual frequency profile of causal flow between nodes (brain areas) was fully known *a priori* by mathematical construction, allowing for simple direct testing. As a result, it was shown that the gPDC misleads the frequency spectral characteristics of causal information flow, such as the location of all local maxima, the correct number of local optima or the decaying behavior of the curve tail at large frequencies, as well as the corresponding strength value of main connections (normalized between 0 and 1, such as iCOH). In contrast, spectral profile calculated by iCOH did not exhibit any of the previous pitfalls rendered by gPDC, in fact, it gives optimal results, i.e., it is able to correctly reproduce the actual spectral properties of the computational toy model. As a consequence, we can assert that iCOH reveals itself as the best possible election for effective connectivity studies.

### 4.5.2. Theoretical aspects

As cautionary aspects relevant to the general topic of functional and effective connectivity, some important considerations prone to bias all calculated results if not properly taken into account need to be briefly explained, in order to avoid spurious connectivity patterns and consequent conclusions (Blinowska, 2011, Kaminski and Blinowska, 2014). Firstly, any connectivity model that does not explicitly get rid of potential instantaneous effects (also called ephaptic conduction) will be systematically flawed, as realistic biological connections are always temporally lagged between the putative causes and effects (Weiss et al., 2013). Consequently, the version of iCOH used in this study is the lagged iCOH, which from the beginning in the MVAR modelling, avoids introducing any term representing instantaneous relationships between sources and sinks (Faes et al., 2010, Pascual-Marqui et al., 2011). Hence, causal flow is assured to be lagged in time, as should be in the proper modelling of a directed synchronization phenomenon between different ROIs (Philiastides and Sajda, 2006, Pascual-Marqui et al., 2014a).

Secondly, the use of an inverse solution method, provided here by LORETA (as MVAR fitting occurs at the level of current vectors associated to brain voxels, and not at recorded EEG channels), instead of conducting effective connectivity analysis at the level of the scalp, is an essential initial consideration that must be adopted, not as an option, but as a requisite to avoid mistaken results (Murphy et al., 2011b). Firstly, it is wrong to extrapolate connectivity results yielded by scalp analyses to the immediate underlying cortical areas (i.e., as resulting from radial projection) (Pascual-Marqui et al., 2011). For instance, a coupling between electrodes O1 and O2 does not generally entail the actual existence of function connectivity between the left and right visual cortices, and furthermore, the question of directionality, as indicated by effective connectivity, gets even more delicate, due to increased complexity in the problem of describing causal arrow flows in a circuit (Stam et al., 2016).

LORETA is a source localization error that has proved to produce zero localization error in the reconstruction of sources of electrical current

generators, a major feature not achieved by any other source localization method (Pascual-Marqui et al., 1994, Pascual-Marqui et al., 2002). Nevertheless, by imposing the realistic condition of smoothness in the solution of the inverse problem for a finitely sampled electromagnetic field (i.e., a small gradient across all directions in a local sphere surrounding a given voxel), the LORETA algorithm is also known to produce limited spatial resolution (especially for deep sources), thus, introducing some blurring in the solution field (Jatoi et al., 2014). It should be insisted that the effect of blurring is a direct consequence of the algorithm provided by LORETA itself, "contaminating" nearby voxels in all directions on a distance range on the order of the resolution width of LORETA voxels (5 mm in cortical gray matter, although a larger blurring range can be expected for deep sources).

Thirdly, the importance of the particular inverse solution method is relevant to the problem of connectivity analyses, given that it exerts a subtle effect on it, described here in the context of using LORETA (Murphy et al., 2011b, Barrett et al., 2012). As it is well-known, the EEG recording technique introduces the systematic error of volume conduction originated from electrical propagation in conductive brain tissue, which will tend to merge information coming from different underlying sources, an unwanted effect that is also incorporated by the inverse solution algorithm (having as input the EEG data) (Vanoosterom, 1991, Bastos and Schoffelen, 2015). As a consequence, estimated cortical signals are instantaneously mixed by volume conduction (typically having a radius of influence of several centimeters, that decreases inversely with distance), which can be problematic in the analyses of high-density EEG, but can be (more or less) dismissed with a low to moderate number of electrodes (27 in our case) (Kaminski and Blinowska, 2014).

In our case, given the blurring introduced by LORETA, an additional mixing of signals in the generated solution can bias effective connectivity errors and distort actual directionality or coupling strength between. For example, two signals corresponding to a given pair of ROIs can be totally uncorrelated at the neuronal level, but due to blurring in the solution, some information might "leak" from one ROI

to the other, hence, any reliable connectivity method would detect a spurious coupling (Pascual-Marqui et al., 2011). We took the caution of defining ROIs separated by considerable distances (minimal centroid ROI distance is 36.4 mm, from the PCC to MCC, thus, sufficient for an effect typical of a 5 mm range between adjacent voxels), so this pitfall introduced by the properties of the LORETA algorithm can be ignored. Besides, as said in the first point, a lagged MVAR model corrects, by construction, the potential caveat of instantaneous effects even for very proximal ROIs, which is also used.

### 4.5.3. Statistics

In this study, we consistently used *t* values, although we do not presuppose nor require that the data follows an exact Student's *t* distribution. Instead, we use the Student's *t* expression as a convenient way to express statistical distance between compared conditions. As iCOH exhibits a frequency dependency, there are 128 tests per directed connection, a number resulting from considering the frequency resolution (0.5 Hz) and the last spectral frequency (64 Hz). Consequently, a considerable number of statistical tests must be performed bin-wise for each connection, which in turn must be repeated 81 times to account for all connections (given the number of ROIs, 9), thus, demanding statistical correction for multiple comparisons, which was addressed by the method of non-parametric randomization of the maximum *t* value across frequency bins (Nichols and Holmes, 2002, Pascual-Marqui et al., 2014a). The plots representing statistics in selected frequency bands (delta and sigma) are also based in *t* tests corrected for multiple comparisons.

Additionally, to tackle the problem of multiple comparisons, we used the statistical non-parametric mapping (SnPM) method. The SnPM is directly built from the empirical distribution of the data itself; therefore, it can be equally applied either to connectivity values between pairs of nodes or single channel measures (e.g., power spectral density). The SnPM method creates a vast number of surrogate permutations, typically as many as several thousand for best statistical accuracy (5,000 in our case), rendering a histogram which provides a statistical threshold. It can be proved mathematically that this method

automatically corrects for multiple comparisons (Nichols and Holmes, 2002).

### 4.5.4. Extended discussion on additional frequency bands

The following discussion expands in more detail on Fig. 4.5 (devoted to the alpha, theta and beta bands). The upper row represents the contrast between after and before SO in the baseline condition. On the left column, corresponding to the theta band, there are 23 significant changes in effective connectivity after SO, as assessed by iCOH, generally conveying a strengthening of the connections. Thus, the MPFC increases flow to the LIPL. The MCC raises flow to the PCC, but decreases it to the RIPL. The PCC increases flow to the MPFC, bilateral inferior parietal lobule and LH. The LIPL reduces flow to the MCC and LH, but increases it to the RIPL. The RIPL raises flow to the MPFC, PCC and LIPL. The bilateral dorsolateral prefrontal cortex reduces flow to the RH. Finally, the LH increases flow to the MCC, PCC and LIPL, while the RH raises flow to the MPFC, MCC, PCC, LIPL and LDPLFC.

The middle row represents the contrast between after and before SO for the recovery condition. On the left column, for the theta band, there are 14 significant connectivity changes. The MCC increases flow to the RIPL. The PCC raises flow to the MCC, bilateral inferior parietal lobule, RDLPFC and RH. The RIPL decreases flow to the MPFC and RDLPFC. The bilateral dorsolateral prefrontal cortex increases flow to the MPFC, while the RDLPFC decreases flow to the MCC and LDLPFC. Finally, the LH increases flow to the MPFC and LDLPFC.

The lower row represents the contrast between recovery and baseline conditions. In the theta band there are 22 significant changes in pairs of ROIs, generally representing decreases in effective connective in the recovery condition as compared to the baseline case (with only two exceptions). Thus, following sleep deprivation, the MPFC shows decreased flow to the MCC, PCC, LIPL and LH. The PCC raises flow to the RILPL, being one of the only two significant increases in the recovery condition compared to baseline. The LIPL lessens flow to the RIPL, RDLPFC and LH, whereas the RIPL reduces flow to the PCC, LIPL and bilateral hippocampus. The LDLPFC decreases flow to the

MCC, PCC, bilateral inferior parietal lobule and hippocampal complex. On the other hand, the RDLPFC reduces flow to the MCC and PCC. Finally, the LH diminishes flow to the PCC, while the RH increases flow to the RIPL (the second exception to flow increase during recovery in the theta band).

The upper row of the middle column shows the contrast between after and before SO in the alpha band for the baseline condition. There are 18 significant effective connectivity changes following the SO transition, which indicate an overall increase. The MPFC increases flow to the PCC, LIPL, LDLPFC and RH. The PCC raises flow to the MPFC, MCC, bilateral inferior parietal lobule, bilateral dorsolateral prefrontal cortex and RH. The bilateral dorsolateral prefrontal cortex increases flow to the MPFC and MCC, with the LDLPFC additionally increasing flow to the PCC. Finally, the bilateral hippocampus increases flow to the LIPL.

The middle row of the middle column represents the contrast between after and before SO in the alpha band for the recovery condition. There are 24 significant changes in effective connectivity following the SO transition, generally indicating a global increase. The MPFC increases flow to the bilateral dorsolateral prefrontal cortex and LH. The MCC increases flow to the PCC, RIPL, LDLPFC and LH. The PCC increases flow to the bilateral inferior parietal lobule, RDLPFC and RH. The LIPL raises flow to the PCC and RH, while the RIPL decreases flow to the RDLPFC and RH. The LDLPFC increases flow to the MPFC, MCC and RDLPFC, while the RDLPFC increases flow to the MPFC, but decreases it to the PCC. Finally, the LH increases flow to the MPFC, RIPL, LDLPFC and RH.

The lower row of the middle column represents the contrast between conditions in the alpha band. There are 26 significant changes in iCOH values when statistically comparing both sleep conditions, in general, indicating a decrease in effective connectivity under sleep deprivation. The MPFC lessens flow to the PCC, but increases it to the LDLPFC. The MCC increases flow both to the MPFC and LIPL. The PCC raises flow to the RIPL. The LIPL reduces flow to the MCC, RILPL, RDLPFC and bilateral hippocampus. On the other hand, the RIPL

decreases flow to the PCC, LIPL, RDLPFC and bilateral hippocampus. The LDLPFC decreases flow to the MPFC, PCC, RIPL and bilateral hippocampus, while the RDLPFC diminishes flow to the MCC, PCC and RIPL. Finally, the LH reduces flow to the PCC, yet, the RH increases flow to the bilateral inferior parietal lobule.

The upper row of the right column represents the statistical contrast between after and before SO in the beta band for the baseline condition. There are 14 significant changes following the SO transition. The MCC reduces flow to the LIPL, LDLPFC and LH. The PCC increases flow to the MPFC, RIPL and RH. The LIPL increases flow to the MCC, RDLPFC and RH. The RIPL increases flow to the LIPL. The LDLPFC increases flow to the MPFC and PCC. Finally, the RH decreases flow to the MPFC and RDLPFC.

The middle row of the right column represents the statistical contrast between after and before SO in the beta band for the recovery condition. There are 13 significant connectivity changes following the SO transition. The MPFC increases flow to the LH. The MCC increases flow to the PCC. The PCC increases flow to the RH. The LIPL increases flow to the MCC and RH. The LDLPFC decreases flow to the PCC and LIPL, but increases it to the RDLPFC. The RDLPFC increases flow to the MPFC, but decreases it to the MCC and PCC. Finally, the LH increases flow to the RH, whereas the RH decreases flow to the MCC.

The lower row of the right column shows the statistical contrast between conditions in the beta band. There are 32 significant changes in effective connectivity between conditions, which convey a general decrease of connectivity strength following sleep deprivation (with only one exception). Thus, after sleep deprivation, MPFC attenuates flow to the MCC, PCC and bilateral inferior parietal lobule. The MCC decreases flow to the RDLPFC. The PCC increases flow to the RIPL (being the only statistical increase), while reduces flow to the bilateral dorsolateral prefrontal cortex. The LIPL diminishes flow to the PCC, RDLPFC and LH. On the other hand, the RIPL decreases flow to the bilateral dorsolateral prefrontal cortex. The LDPFC reduces flow to the MCC, PCC, bilateral inferior parietal lobule and bilateral hippocampus. The RDLPFC decreases flow to the same areas than its

left counterpart, except for the RH. Finally, the LH reduces flow to the MCC, PCC, bilateral inferior parietal lobule and LDLPFC, while the RH lessens flow to the MCC, PCC and RDLPFC.

# 5. Summary of connectivity methods and their applications in sleep research

## Abstract

A selection of functional and effectivity connectivity methods designed to explore network patterns in EEG/MEG data is exposed to foster its use and divulge the underlying theory, with a predominant emphasis on sleep research and previous applications in this field. We aim to describe the potential strengths as well caveats associated to each method in simple terms, and how they are generally interrelated (e.g., by establishing a hierarchy of limitations or range of application). The review we describe encompass both linear (e.g., Granger causality) and non-linear methods (e.g., mutual information). Examples in the sleep neuroscience help to illustrate the main theoretical aspects characterizing the methods, and how fruitful they reveal for future research on this area, or more generally, in either functional or effective connectivity studies. Special attention is given to explain the differences in the concepts of functional and effective connectivity, that may still be sometimes confused, and the particular interpretation of the causality concept in the context of neuroscience (e.g., as directed information flow between brain regions). No advance knowledge in mathematics is required to understand the explanation of the different techniques, which are written in an intuitive and direct way.

**Keywords:** brain connectivity, functional connectivity, effective connectivity, synchronization, directionality, applied mathematics, sleep.

## 5.1. Introduction

*"Without mathematics, there's nothing you can do. Everything around you is mathematics. Everything around you is numbers."* Shakuntala Devi[5.1]

In this review, we discuss a collection of methods to investigate functional and effective connectivity analyses by means of EEG/MEG data. Although the connectivity methods themselves are completely general in their formulation, and the potential range of application is much broader than the one presented here, we focused on applications to sleep research (especially, in humans), and aimed to provide a general understanding of these methods for researchers in this field. Consequently, the general structure is a summary of connectivity methods followed by particular applications in the context of sleep (especially in, although not exclusively restricted to, humans). When it was not possible to provide examples with sleep recordings, wake recordings were used instead (e.g., comparing eyes open versus eyes closed).

When discussing connectivity patterns of the brain, we have to dissociate between structural (anatomical), functional and effective connectivity (Friston, 1994, Sporns et al., 2007, Friston, 2011, Reimann et al., 2017). Brain connectivity can be described at various levels of scales (Sporns et al., 2007, Dipasquale and Cercignani, 2016, Houck et al., 2017). In the following, we focus on the macroscale, as we discuss applications to EEG recordings reflecting large numbers of neurons or neuronal populations in order to identify brain regions that are interconnected.

The methods discussed in this review are among the most well-known and validated techniques to tackle the connectivity problem in the context of EEG/MEG data. In the first place, information theory has been selected because it constitutes a rich body of techniques that are able to cover multiple levels of connectivity investigations practically simultaneously (all methods share a common root, their relationship is similar to the multiple branches of a tree) (Ramanand et al., 2010a, Chan et al., 2013, Kaminski et al., 2001). Consequently, information theory has become popular since it incorporates efficiently and

elegantly the following aspects: the strength of connections (including a possible non-linear character), the main directionality, and potential effects of confounding variables hidden inside the network structure (Vicente et al., 2011, Pires and Perdigao, 2012, Wollstadt et al., 2014). A pertinent remark related to information theory concerns the expression “entropy” that appears many times (e.g., Shannon entropy, Renyi entropy, Transfer entropy) and may lead to confusion. Thus, one might assume that “approximate entropy” and “sample entropy” can also be employed for connectivity studies. However, this is not the case: both approximate and sample entropy are methods established for analysis of a single channel (basically, they are used to quantify signal unpredictability) (Fonseca et al., 2015). Consequently, they have not been included in this review that focuses on the connectivity within a network, and not on univariate measures (isolated channels, ignoring mutual influences).

The next selected method is Granger causality (G-causality) (Granger, 1969, Friston et al., 2014). Although this method was originally invented to be applied in a very different field (economics and prediction of stock market trends), during the last two decades it has gained considerable attention among neuroscientists as well, who also find it useful for the study of brain connectivity (Kaminski et al., 2001, Faes and Nollo, 2013, Seth et al., 2015). The reason for this emergent interest can be attributed to the capacity of Granger causality to distinguish the dominant variable of an interacting pair of nodes (both in the time and the frequency representations), the simplicity of its mathematical formulation (due to its linear character) and the easiness of computational implementation. Although it reveals problematic when applied to combinations of multiple interacting channels, getting worse the bigger is the size of the network (Barrett et al., 2010, Faes et al., 2014, Seth et al., 2015).

Thus, in order to overcome the problems arising when Granger causality is applied to networks of moderate to high complexity, new connectivity methods circumventing these obstacles were developed: the Direct Transfer Function (DTF), the Partial Directed Coherence (PDC) and the isolated Effective Coherence (iCOH). All these methods

share in common important aspects: a linear formulation (expressed, in every case, in the frequency domain), being able to include the whole network in a single analysis (i.e., not requiring multiple applications of pairwise analyses, that would likely not lead to reliable conclusions) and a normalization ranging from a minimum of 0 (indicating no synchronization) to a maximum of 1 (for full synchronization) (Astolfi et al., 2007, Silfverhuth et al., 2012, Pascual-Marqui et al., 2014a, Bonhomme et al., 2016).

Additional methods addressing either functional or effective connectivity are used in computational neuroscience (not covered in this review). Some of the most interesting approaches are: phase-slope index (PSI), dynamic causal modeling (DCM), structural equation modeling (SEM), triplet synchronization index, etc. (Friston, 2011, Ewald et al., 2013, Kralemann et al., 2014, Bastos and Schoffelen, 2015).

In the applications of all these methods, a systematic problem that arises is the choice of the EEG reference (Liu et al., 2015). The problem of the selected reference is exclusively related to the EEG (as a potential difference is measured) and does not appear for MEG data (Oliviero et al., 2004, Houck et al., 2017, Piano et al., 2017). Consequently, the selected reference will have some impact on the final connectivity indices and as well in other univariate measures (e.g., -to some degree- in power spectral density) (Cohen, 2015, Kuo et al., 2016). A common recommendation is to avoid the average reference, or at least be cautious with the fact that it may render spurious connections, due to the following reasons: firstly, because it may transfer activity from highly active to moderately resting channels (therefore, modifying the topographic properties of the network, e.g., power maps, but also graph analyses), and secondly, as the average reference entangles or mixes the network signals (i.e., subtracting the average of all signals) that not only correlation claims, but also directionality of causal flows may be ultimately unreliable (Kaminski et al., 2001, Kaminski and Blinowska, 2014). Other common problems relate to the choice for the sampling rate (sufficiently high to avoid aliasing, but not enough so that fast noise fluctuations begin to being included in the subsequent calculations), the embedding dimension for

state space representations (traditionally, by following Taken's theorem) or the window choice (e.g., Gaussian, Hanning or simply no window) (Carmeli et al., 2005, Seneviratne, 2014, Shine et al., 2015). The potential consequence of not following these simple advices in the preprocessing stage is that, even when the connectivity techniques are rightfully applied, the final results obtained might be biased or wrong. In this review, we focus on two types of connectivity: functional connectivity (FC) refers to the statistical dependence or correlation between distinct neuroanatomical brain regions, while effective connectivity (EC) refers to the influence that one neural system exerts on another one with regard to causal interactions, including the directionality of the interaction (Friston, 2011, Stam et al., 2016, Reimann et al., 2017). Mathematically, in a given neuronal network (as e.g. measured through EEG recordings), both FC and EC values are represented by means of a matrix, whose order or size correspond to the number of channels employed (Faes et al., 2012, Faes and Nollo, 2013, Stam et al., 2016). This matrix is symmetrical for FC analysis, since correlation coefficients are not sensitive to the direction of propagation (Kaminski et al., 1997, Jeong et al., 2001, Friston, 2011, Liu and Aviyente, 2012). On the other hand, EC is designed to take into account explicitly the directionality of the interaction, hence, they are represented by completely general matrices, i.e, not requiring any symmetric character (Sporns et al., 2007, Erla et al., 2009, Goldenberg and Galvan, 2015, Huang et al., 2015).

The problem of causality has intrigued mathematicians, natural scientists and philosophers throughout history (Hlavackova-Schindler et al., 2007, Wibral et al., 2014). Skeptical philosopher David Hume (1711-1776) argued that every causal claim is ultimately based on experience, and experience similarly based on the assumption that the future is a consequence of the state of the past, which is based again on induction by experience, leading to a circular logic. Hence, a part of Western philosophy has challenged our basis to assert a true causal link on natural phenomena (as a consequence, Hume only admitted correlation combined with induction, at best). As a second example, econometricians and statisticians prefer to say the variable *X* "Granger-causes" variable *Y*, rather than directly employing the word "causes"

(in this case, because Granger causality –to be described later in this review- does not account for the possible existence of latent or confounding variables) (Granger, 1969, Friston et al., 2014, Seth et al., 2015).

Additionally, the *post hoc ergo propter hoc* fallacy (Latin for "after this, therefore because of this") asserts that just because an event *A* precedes an event *B* in time, event *A* is falsely assumed to actually cause event *B*. This is a commonly committed fallacy of the layperson. The following statement could be an example of this fallacy: “after taking a particular drug, a patient improves in the recovery from a disease, therefore it is concluded that the drug helped to heal that patient”; this statement can be questioned if there is no control condition to compare with.

The classical statement “correlation does not imply causation” is well known and should not be ignored. Because the methods commonly used in applications of functional connectivity (such as cross-correlation in the time domain, or coherence, in the frequency domain), are only able to establish the existence of correlations in the variation between two variables, purely correlational methods are not appropriate to investigate causal relationships, although they can bring some indirect and partial clues on it (Schelter et al., 2006, Frenzel and Pompe, 2007, Silfverhuth et al., 2012, Sommerlade et al., 2012). If the first question refers to finding out whether two signals are statistically correlated, the second question must be: is there a causal relationship, or could other latent variables explain the assessed coupling? Furthermore, if the link is really causal, as latent variables are ruled out, is it possible to determine a predominant direction of transmission? (Sommerlade et al., 2011, Kralemann et al., 2014, Pascual-Marqui et al., 2014a, Seth et al., 2015).

Consequently, causality is addressed in the context of computational neuroscience by the introduction of effective connectivity techniques (Friston, 2011, Liu and Aviyente, 2012, Kuhn et al., 2014, Stam et al., 2016). In general, effective connectivity methods do not absolutely solve the issue of latent variables, but the most popular among them (when performing an overall analysis of all channels simultaneously, most commonly linear, as explained later) may be extremely powerful

in its detection, offering an excellent reliability (Korzeniewska et al., 2003, Khadem and Hossein-Zadeh, 2014, Huang et al., 2015). In this way, causality can be formally investigated in computational neuroscience with the introduction of effective connectivity techniques, ranging from the very simple (e.g., Granger causality, very common, but flawed to some degree by false positive connections) to the most complex approaches (e.g., transfer entropy, a non-linear method derived from information theory) (Vicente et al., 2011, Barrett et al., 2012, Friston et al., 2014, Pascual-Marqui et al., 2014a).

The use of surrogates is a suitable method for the statistical evaluation of either functional or effective connectivity (Kunhimangalam et al., 2008, Faes et al., 2009, Kralemann et al., 2014). The reason for introducing surrogates is due to a systematic bias, which is unavoidable in practice, i.e., the finiteness of the data length in an actual record (Palus, 1997). Because epochs to be analyzed are always finite, a systematic error is implicitly present in the results of connectivity analyses (Le Van Quyen et al., 2001, Schelter et al., 2006). For example, even for two completely unrelated white noise signals, the finiteness of the epoch would render a value greater than zero (although most probably small), but not exactly zero (Jeong et al., 2001). That opens the question on how to validate connectivity results, in order to be sure that connectivity values obtained are reflecting possibly small, yet significant connectivity values, as opposed to no connection at all (Baccala and Sameshima, 1999, Roach and Mathalon, 2008). Furthermore, the existence of random noise fluctuations over time also creates a contributing factor reducing the statistical distance between significant and non-significant connections, so both type 1 and 2 errors are possible (especially the latter, similar to lack of enough statistical power due to noise) (Kralemann et al., 2014).

Thus, surrogates are a popular method to overcome the problem of the previously explained systematic error, aiming to test for statistical significance (Kunhimangalam et al., 2008, Faes et al., 2009). In this approach, the value of connectivity is tested for statistical significance under the null hypothesis of no significance, that is, the connection value is not strong enough to have any real effect (Nichols and

Holmes, 2002). Surrogates of a given signal preserve the same data length and other basic statistical properties of the signal (most notably, the autocorrelation function or power density spectrum) (Kunhimangalam et al., 2008). However, they are typically modified either in the time domain (called independent identically distributed surrogates, formed by random sample shuffling, hence maintaining the same mean, variance and histogram distribution) or in the frequency domain (called Fourier transform surrogates, permuting the phases, but without changing the amplitudes, thus, keeping the same power density spectra) (Faes et al., 2009). In practice, the number of possible surrogates that can be formed with EEG/MEG signals is huge, as typically, epochs are composed of several seconds, sampled at a minimum of 100 Hz. The procedure of creating surrogates is supposed to destroy the a priori correlations that may exist between a pair of connected channels, because when at least one signal has been disorganized, the coupling is theoretically reduced as the remaining statistical significance that could exist is mostly eliminated (Faes et al., 2009). When this procedure is repeated many times (typically, hundreds of times), so that signal 1 remains the same, but signal 2 is replaced by a surrogate, a histogram with these values can be built (Palus, 1997). If the value of connectivity for the unmodified pair of signals is greater than a given threshold in the histogram, the connection can be considered to be statistically significant (Le Van Quyen et al., 2001). In general, the type 1 error is inversely proportional to the number of surrogates used (e.g., $\alpha$=1% for approximately a hundred surrogates), thus, aside from the computational load, a greater number of surrogates will improve the statistical validation of analyses (Astolfi et al., 2007). In neuroscience, the problem of multiple testing appears frequently (typically, when including a large number of voxels); in this context, methods such as bootstrapping and statistical non-parametric mapping (SnPM) allow to address multiple comparisons by using surrogates without requiring to presuppose any a priori probability distribution, thus, being flexible to the structure of the particular empirical distribution (Nichols and Holmes, 2002, Kunhimangalam et al., 2008, Faes et al., 2009).

## 5.2. Functional connectivity (FC)

### 5.2.1. Information theory

Information theory is an important branch of signal analysis that has encountered applications in a wide variety of scientific fields, including computational neuroscience (Jeong et al., 2001, Vicente et al., 2011, Melia et al., 2015). The concept of information was first introduced by the mathematician C. E. Shannon in 1948 (Shannon, 1948). In simple terms, the content of information (also called Shannon entropy) of a general signal can be defined as the average number of bits needed for its codification (Shannon, 1948, Barrett et al., 2012, Wollstadt et al., 2014). An alternative but equivalent interpretation describes information as a variable to quantify the total uncertainty of a system, preceding any measurement on it (since measuring would imply extraction of information about the system, hence, lowering the amount of uncertainty) (Escudero et al., 2009, Kinney and Atwal, 2014).

Information theory presents a potent set of tools that are worthy to consider in studies related to either functional or effective connectivity analyses, such as mutual information, conditional entropy, transfer entropy (Escudero et al., 2009, Lee et al., 2012, Melia et al., 2015). In the context of effective connectivity, the most important index in information theory is transfer entropy (TE) (Hlavackova-Schindler et al., 2007, Khadem and Hossein-Zadeh, 2014) and will be discussed later.

### 5.2.2. Mutual information (MI)

A crucial concept in information theory regarding functional connectivity studies is mutual information (MI; Table 5.1) (Shannon, 1948, Hlavackova-Schindler et al., 2007, Ramanand et al., 2010a). There exist several yet equivalent ways of defining MI. Probably the most straightforward definition is that MI measures the reduction of uncertainty of a signal (as indicated by calculation of Shannon entropy), given acquisition of knowledge (i.e., information) of another signal (Jeong et al., 2001, Julitta et al., 2011, Chan et al., 2013). It should be noted that the second signal can be either a completely different signal, or past values (contained in a certain number of epochs preceding the current epoch) belonging to the same original signal (Julitta et al., 2011, Melia et al., 2015). An alternative definition is that MI represents the excess of total information of a system, under the assumption (regardless of whether correct or not) of mutual independence of its components (Shannon, 1948, Le Van Quyen et al., 2001, Pires and Perdigao, 2012). This second definition may be better understood using a Venn diagram (Fig. 5.1). The key concept is that MI is a variable used to quantify statistical dependencies within or between pairs of signals in a very general way, as it captures and combines different classes of potentially existing correlations (Le Van Quyen et al., 2001, Escudero et al., 2009, Kinney and Atwal, 2014). In particular, as a very important characteristic, MI is also sensitive to non-linear correlations (whereas linear methods are inherently limited in this regard), as well as other types of correlations, e.g., cross-frequency transmission between different frequency bands (Jeong et al., 2001, Escudero et al., 2009, Vicente et al., 2011).

Remarkably, MI is symmetrical, meaning that the order of consideration of the signals is completely interchangeable; this property is in agreement with MI being regarded as a method of functional connectivity (Frenzel and Pompe, 2007, Melia et al., 2015). One important theorem MI satisfies is that when applied to two independent signals, MI reaches theoretically zero, as obtaining information of one system does not reduce the uncertainty of the other one (e.g., the electrocardiogram of an individual generally does not help to predict the state of the electrooculogram) (Ramanand et al.,

2010a). Alternatively, using the second definition of MI, the excess of information measured would also be zero (Jeong et al., 2001). Using the mathematical definition of MI, it can be demonstrated that the previous theorem is a direct consequence of the definition of independent events in probability theory, where the joint probability distribution factorizes into its independent components (Shannon, 1948, Hlavackova-Schindler et al., 2007). Nevertheless, in practice, the length of data cannot be infinite but is always finite, thus yielding MI values greater than zero, even for white noise (i.e., theoretically completely uncorrelated processes) (Pires and Perdigao, 2012). As always, the use of surrogates (see introduction) or other type of statistical criterion is imperative in conjunction with the application of any connectivity method (Faes et al., 2009, Ramanand et al., 2010a).

Another property of MI is that, when applied to EEG/MEG signals, MI reflects the total statistical dependence between derivations (MI can be generalized to be applied to any number of signals, not just a pair, yet that is the most common case), which using a Fourier decomposition, can be divided into two parts: correlation of the phases (commonly referred to as synchronization) and of the amplitudes (or power) (Palus, 1997, Le Van Quyen et al., 2001, Bruns, 2004). If the interest is restricted to the synchronization of phases in a certain frequency range, discarding possible correlations of the amplitudes, the analyses should consider, not the raw signals, but the extraction of the associated phases as a function of time (Palus, 1997, Le Van Quyen et al., 2001, Kralemann et al., 2014). In order to obtain the temporal evolution of the phases, either the Hilbert transform, the wavelet transform, or the short time Fourier transform are suitable methods (Bruns, 2004, van Vugt et al., 2007). Conversely, it is also possible to study statistical dependencies in amplitude or power, if time-series representing the evolution of power spectral density as a function of time are used instead (Wright et al., 1995, Bruns, 2004). Therefore, MI applied to the original raw signals actually combines both aspects (that is, phase and amplitude dependencies) simultaneously, but in order to know specifically the weight of contribution to the final values, separate analyses in the way described above are recommended (Le Van Quyen et al., 2001, Na et al., 2006).

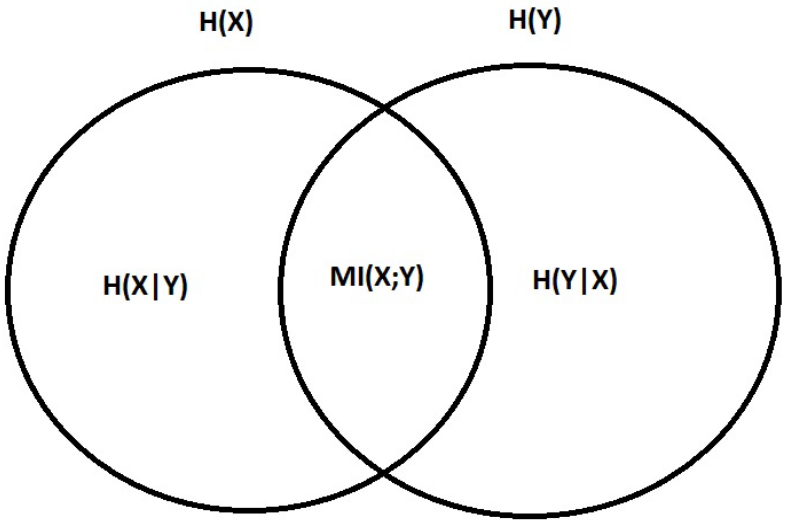


***Fig. 5.1.*** *Venn diagram graphically illustrating Mutual Information, MI(X;Y). MI between signals X and Y is given by the intersection of the Shannon entropies H(X) and H(Y).*

#### 5.2.2.1. Cross-mutual information (CMI)

MI can be applied either to completely different signals, or to different versions of the same original signal (but involving a time shift) (Escudero et al., 2009, Julitta et al., 2011). This distinction will be explained in more detail in the following. Cross-mutual information (CMI; Table 5.1) is the term used when MI is explicitly applied to two different time series (e.g., two different EEG/MEG derivations), and is used to quantify the content of information shared between them, measured in bits (Na et al., 2006, Chan et al., 2013, Melia et al., 2015). It can be considered as a generalization of cross-correlation (in the time domain) or coherence (if expressed in the frequency domain), and includes non-linear dependencies besides the linear correlations (Jeong et al., 2001, Melia et al., 2015). This measure is classified as a functional connectivity technique, because it does not tell which channel is transmitting information to the other one (in mathematical terms, CMI is symmetrical when permuting the signals), but reflects the existence of a correlational –not necessarily causal- relationship (Ramanand et al., 2010b, Bastos and Schoffelen, 2015). Nonetheless, estimation of the main directionality of interaction can be achieved by introducing a time lag in one of the signals, and then observing if the

maximum of CMI appears either for positive or negative values of the time lag (Jeong et al., 2001, Frenzel and Pompe, 2007). However, in order to study the directionality problem in the context of information theory in a fully rigorous way, the most recommended variable, as indicated above, is transfer entropy (Vicente et al., 2011, Lee et al., 2012, Silfverhuth et al., 2012, Wollstadt et al., 2014).

#### 5.2.2.2. Auto-mutual information (AMI)

On the other hand, when MI is computed using just a single time series, it receives the name of auto-mutual information (AMI; Table 5.1) (Jeong et al., 2001, Na et al., 2006, Melia et al., 2015). AMI can be understood as a measure proportional to the degree in which past values of a signal are useful to predict its current value, or future evolution (Jeong et al., 2001, Julitta et al., 2011). Mathematically, AMI just introduces a time lagged version of the original signal (Escudero et al., 2009, Julitta et al., 2011). In this way, AMI becomes a function of a temporal parameter, pretty much in parallel with the classic and well-established auto-correlation function, its linear counterpart (e.g., both functions are periodical when applied to periodic signals and generally decay asymptotically towards zero when raising sufficiently the time lag) (Ramanand et al., 2010b, Silfverhuth et al., 2012). Additionally, the maximum of AMI is located at zero lag (as for the auto-correlation function), where it corresponds exactly to the Shannon entropy of the signal (Escudero et al., 2009, Ramanand et al., 2010b). However, if AMI is compared to the auto-correlation function, two important differences must be stressed: first, whereas the auto-correlation function can assume either positive or negative values, AMI is always positive (by definition, as derived from MI) (Jeong et al., 2001, Julitta et al., 2011). Second, the auto-correlation function is a linear estimation, while AMI is completely general (i.e., it can be safely applied to highly non-linear signals, where it would outperform the auto-correlation function in the detection of complex trends) (Julitta et al., 2011).

The lag parameter of AMI plays the role of a “memory”. Thus, when comparing two different signals, the one that decays quicker (e.g., as

characterized by reaching the first minimum earlier) loses the causal influence given by its past values faster, i.e., showing higher unpredictability (Ramanand et al., 2010a, Melia et al., 2015). Consequently, a shorter lag until the first minimum of AMI indicates that the temporal evolution of the signal is more short-term oriented, being difficult to forecast future values except for very proximal instants (Ramanand et al., 2010b). These features are associated with an increased signal complexity and hence have been proposed to serve as a proxy for neuronal network complexity in the context of neuroscience (Escudero et al., 2009, Seth et al., 2011). Because the latter is a property specifically of non-linear and chaotic systems (that is, forecasting is more complicated, and past causal influences are diluted at a faster pace), AMI can be used to predict the "entropy" of a signal (Pires and Perdigao, 2012, Huang et al., 2015). Shorter decay times indicate high unpredictability and complexity of the signal, possibly reflecting the underlying presence of non-linear dynamics (Ramanand et al., 2010a, Ramanand et al., 2010b, Huang et al., 2015).

Finally, practical aspects for computing all mutual information related variables (CMI, AMI, etc.) must be considered. The classical approach of computing Shannon entropy and mutual information indexes uses histograms to estimate probability densities (Jeong et al., 2001, Escudero et al., 2009). A major caveat is that there is no unique way to generate a histogram in order to calculate the probability distribution (Lee et al., 2012). Different yet comparable number of bins would modify the resulting distributions, hence affecting the final mutual information or entropy values (the same problem occurs for other related measures, such as transfer entropy) (Hlavackova-Schindler et al., 2007, Wibral et al., 2014). Hence, when taking a too coarse histogram with a small number of bins, the histogram tends to flatten, consequently, MI would be underestimated (Escudero et al., 2009). Reciprocally, a too small bin width would bias the probability distribution, causing the opposite effect, therefore, overestimating MI (Escudero et al., 2009, Barrett et al., 2010). Different authors give their own criterion to choose an optimal value for the number of bins (Tass et al., 1998, Le Van Quyen et al., 2001, Kinney and Atwal, 2014), but as a rule of thumb to select a reasonable number of bins, one should take approximately the square root of the number of samples (duration

of the epoch times the sample frequency) of the epoch analyzed (Le Van Quyen et al., 2001).

### 5.2.3. Partial information (PI)

Another relevant measure to introduce is partial information (PI) (Frenzel and Pompe, 2007, May et al., 2008). PI is a variable that depends on two signals; it quantifies the remaining uncertainty to describe the state of a given signal when knowledge of another signal is available (Faes et al., 2012, Wollstadt et al., 2014). Mathematically, it is defined in a similar way as the expression for Shannon entropy, but using a conditional probability instead (Shannon, 1948). It is immediate to recognize from the definition of PI a connection with MI; indeed, MI between two signals $X$ and $Y$ can be recovered as Shannon entropy of one of the signals (e.g., $X$) after subtraction of the corresponding PI of the pair, ($X|Y$) (see Fig. 5.1 and 5.2, the latter when generalized for three variables) (Ramanand et al., 2010a). As an example, in an experiment involving the use of the electrooculogram (EOG), if only activity of one of the eyes is recorded (or used), the PI implies a reduction in the a priori uncertainty for the activity of the missing eye, as the presence of movements in one eye correlates with the state of the missing eye (e.g., during rapid eye movements of REM sleep). Similarly, mutual information would indicate in this example the average number of bits that the uncertainty about the state of the missing eye has been reduced, given the information provided by the recordings of the other eye.

#### 5.2.3.1. Partial mutual information (PMI)

Finally, from partial information, a more complex but valuable variable can be directly derived, the partial mutual information (PMI; Table 5.1), also called conditional mutual information (Hlavackova-Schindler et al., 2007, May et al., 2008). Whereas PI depends on only two variables or signals, partial mutual information includes a minimum of three variables (May et al., 2008). In order to motivate its application, suppose that for two main variables *X* and *Y* (representing signals extracted from two different channels or regions of interest in the brain), the application of MI has rendered a high value, that in addition has been statistically validated (for instance, by means of surrogates, surpassing the cut-off for significance in their distribution). However, the correlation found does not truly guarantee the existence of a direct coupling between *X* and *Y*, as confounding variables (denoted by *Z*), could be in fact the responsible agent in the coupling (e.g., *Z* is a major hub in the brain networks, driving the synchronization of *X* and *Y* simultaneously). Thus, MI answers the question of detecting statistical relationships between different channels (even better than other methods such as coherence, because it also takes into account non-linear couplings), but is not able to tell if the nature of the link is direct or caused by confounding variables (Frenzel and Pompe, 2007, Chan et al., 2013). In order to assure that the observed link reflects an actual interaction, PMI between *X* and *Y* with respect to *Z* could be employed, and compared with the original value of MI between *X* and *Y*. Then, if the values obtained in both cases were the same (i.e., a non-significant statistical difference), it would be indicative of no involvement of *Z* (the opposite is also true). The latter is the same procedure also followed with partial coherence (a linear counterpart of partial mutual information, where classical coherence assumes the role of mutual information) (Astolfi et al., 2007, Frenzel and Pompe, 2007).

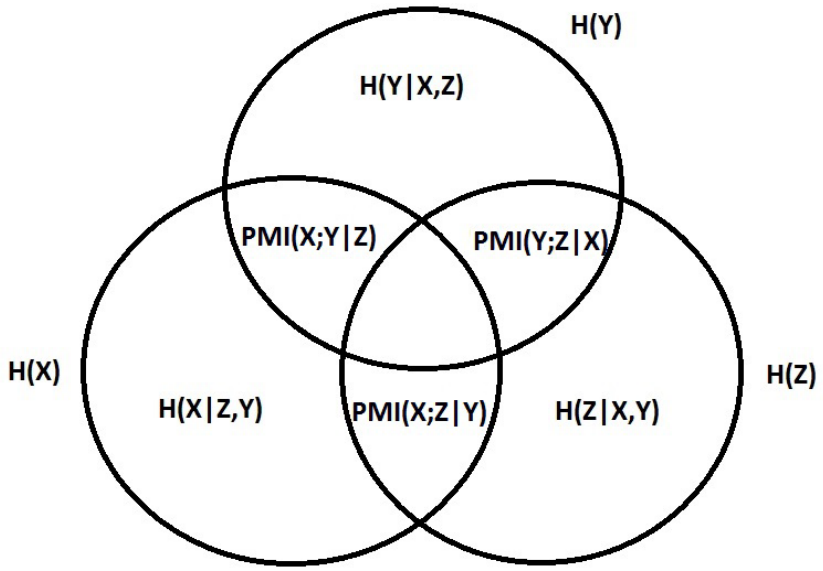


***Fig. 5.2.*** *Venn diagram illustrating partial mutual information between X and Y, given Z, PMI(X;Y|Z) in the figure.*

Basic properties of the PMI are the following: as can be deduced from Fig. 5.2, PMI is either equal or smaller than MI (for example, in Fig. 5.2 *PMI(X;Y|Z)* is either equal or smaller than the intersection between *H(X)* and *H(Y)*, that would correspond to *MI(X;Y)*). A relevant theorem is that if PMI reaches its theoretical inferior limit (nearly zero) and yet MI is significantly greater than zero, this result entails that variable *Z* completely reduces the uncertainty for the description of the interactions between channels *X* and *Y*, hence, all the information regarding the interactions among these channels is implicitly included in *Z* (May et al., 2008). Therefore, *Z* is responsible for the interactions between *X* and *Y*: they might be really synchronized, displaying functional connectivity, but there is no causal or effective connectivity among them. Thus, if PMI is not significantly different than MI, then variable *Z* is not acting as a confounding variable but is rather independent, whereas if PMI is significantly reduced, it provides a direct indication of *Z* being directly involved in the observed correlation between *X* and *Y* (Frenzel and Pompe, 2007).

A subtle but important point in the previous discussion should be noted, in order to prevent mislead reasoning about functional connectivity using PMI: the fact that *Z* does not lower sufficiently MI between *X* and *Y* does not prove that *X* and *Y* are necessarily causally related. In this regard, three reasons might be considered to explain this possibility without entailing a contradiction: 1) The channel *Z* considered is not the right confounding variable (it has to be replaced with other brain region(s)). 2) The actual driver acting as a confounding variable was not recorded, thus, it will remain hidden from detection. 3) Instead of a single brain region, well localized, acting as a principal source of synchronization, it could also be a network spread through many cortical areas, acting in an integrated fashion: individually, particular channels *Z* would not appear necessarily to be substantially involved, but integrating its action in an additive form in the context of a network, total contribution of the network could act as an "effective" confounding variable in the connectivity link between *X* and *Y*.

In summary, the main utility of PMI is that it can be used to detect confounding variables affecting functional connectivity, rendering it an extremely valuable tool for connectivity investigations. Moreover, it automatically incorporates the detection of non-linear trends, hence, theoretically outperforming methods created for the same ends such as partial coherence. PMI might be computationally more demanding with an increasing number of channels and mathematically less straightforward to calculate or code.

There are two main ways to estimate PMI. Firstly, directly from the definition, as it can be computed combining the value of Shannon entropy of all the variables involved (the main variables *X* and *Y*, and the presumed confounding *Z*) and the mutual information of the pair of interest (*X* and *Y*) (Le Van Quyen et al., 2001, May et al., 2008). In this approach, the construction of several histograms in order to estimate the probability distribution functions is needed. An alternative method to compute PMI, independent of the somehow arbitrary construction of histograms for Shannon entropy, has been studied by (Frenzel and Pompe, 2007). In this approach, a three-dimensional vector must be constructed, whose components represent the values of *X*, *Y* and *Z* at a

given time. Over time, a cloud of points will hence be produced, each corresponding to a time sample. For a given point, a sphere with radius equal to the distance between this point and its *k* nearest neighbors (with *k* typically from a few up to around 30, but is very robust towards variations) is built (see further details in (Frenzel and Pompe, 2007)). Although the latter method does not depend on a number of bins, it might be difficult to implement and requires more computational power. However, the authors insist on the advantages rendered with this alternative procedure, giving, on the contrary than histograms, an unbiased estimator for PMI and MI, even with a relatively reduced amount of time samples (around 1,000 samples, or 8 s sampled at 125 Hz) (Frenzel and Pompe, 2007).

| **Variable name (Abbreviation)** | **Brief description** |
|---|---|
| Shannon entropy (H) | A measure related to probability theory to quantify the content of information of a signal, commonly expressed in bits. The more random and irregular the signal, the greater its Shannon entropy. $$H(X) = -\sum_{x} P(x) \cdot ln\big(P(x)\big)$$ |
| Cross-mutual information (CMI) | Non-linear functional connectivity measure between a pair of distinct signals, symmetric under permutation. A generalization of the cross-correlation function to include non-linear correlations. $$CMI(X,Y) = \sum_{x,y} P(x,y) \cdot ln\left(\frac{P(x,y)}{P(x)\cdot P(y)}\right)$$ |
| Auto-mutual information (AMI) | Single channel measure, dependent on one signal. It is similar to CMI when using the original signal and a delayed version as the second signal. A generalization of the auto-correlation function to include non-linear correlations, but only assuming positive values. The initial slope of decay of AMI can be used to quantitatively characterize the "complexity" (or unpredictability) of a signal. |

| | |
|---|---|
| | $$AMI\big(X(\tau)\big) = \sum_{x_t, x_{t+\tau}} P(x_t, x_{t+\tau}) \cdot ln\left(\frac{P(x_t, x_{t+\tau})}{P(x_t) \cdot P(x_{t+\tau})}\right)$$ |
| Partial information (PI) | A measure to describe the Shannon entropy of a signal *X* when the part of its dynamics shared with another signal *Y* is removed. If it is not significantly smaller than Shannon entropy of *X*, it implies no correlation between these two signals. It is not symmetrical in its arguments. $$PI(X\|Y) = -\sum_{x,y} P(x,y) \cdot ln\left(\frac{P(x,y)}{P(y)}\right)$$ |
| Partial (or conditional) mutual information (PMI) | A measure to discriminate the existence of possible confounding variable signals between signals previously known to be correlated by CMI. When PMI including a third channel *Z* is similar to CMI between *X* and *Y*, then *Z* is not involved in the coupling; otherwise, it explains partially the coupling due to indirect interactions related with variable *Z*. $$PMI(X,Y\|Z) = \sum_{x,y,z} P(x,y,z) \cdot ln\left(\frac{P(x,y,z) \cdot P(z)}{P(x,z) \cdot P(y,z)}\right)$$ |

| **Variable name (Abbreviation)** | **Brief description** |
|---|---|
| Transfer entropy (TE)<br><br>(a measure of effective connectivity) | Non-linear effective connectivity measure between a pair of signals, non-symmetric in its arguments. A generalization of Granger causality in the context of information theory, to include non-linear causal interactions.<br><br>$TE(X \rightarrow Y) = \sum_{y_{t+1}, y_t, x_t} P(y_{t+1}, y_t, x_t) \cdot ln\left(\frac{P(y_{t+1}|y_t, x_t)}{P(y_{t+1}|y_t)}\right)$ |

***Table 5.1:*** *Brief description of information theory derived functional connectivity and other important measures needed in this context from. The last one (transfer entropy) is a measure of effective connectivity.*

### 5.2.4. Applications to sleep EEG recordings

Ramanand and colleagues applied information theory derived measures (CMI and AMI) to characterize differences in sleep patterns of middle-aged and elderly women (Ramanand et al., 2010a, Ramanand et al., 2010b). The sleep and wake EEG of 15 middle aged (age range: 45-50 years) and 15 elderly women (age range: 71-86 years) was recorded.

Ramanand and colleagues computed the CMI and AMI for derivations C3A2 and C4A1 (30-s epochs in all cases), to search for differences during waking, NREM (S2 and S3) and REM sleep. The following bands were analyzed: delta-low (0.2-2 Hz), delta-high (2-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma (12-16 Hz) and beta (16-24 Hz). Average CMI (A-CMI), average AMI (A-AMI) in the time window of -0.2 to 0.2 s of the CMI and AMI curves, and the decay rate of the AMI curves (all average of C3A2 and C4A1) were determined.

In (Ramanand et al., 2010a), it was proven that CMI was significantly lower in the delta, theta and sigma band for the elderly group compared to the middle-aged group (see Fig. 5.3). According to the authors, this result is indicative of more independent cortical dynamics, or alternatively, less synchronized coupling between distinct cortical sites, which would lead to reduced sleep efficiency and more frequent arousal across the night (explaining the disturbed sleep reported by the elderly, e.g., the greater sleep fragmentation). Additionally, a general result was that with increasing sleep depth, cortical correlation or interdependence (as assessed by CMI) increased monotonically until reaching a maximum in stage 3, with this trend being more clearly evident (in a statistical sense) in the middle-aged group. Finally, auto-mutual information also increased with sleep depth, thus, being highest during slow wave sleep (stage 3), corresponding to a cortical state of high predictability (and therefore, larger auto-mutual information values) due to the low frequency, high amplitude oscillations characteristic of deep sleep. The latter result was further supported by the decreased decay rate with increased sleep depth, as higher decay rate would imply greater unpredictability, and consequently, lower auto-mutual information.

In (Ramanand et al., 2010b), the authors investigated the effects of arousal during sleep in the elderly, showing that after periods following cortical arousals, the strength of cross-mutual information coupling was significantly reduced. This effect was particularly notorious on central derivations (C3 and C4 electrodes, referenced to mastoids). On the other hand, periods preceding arousal had comparatively greater interdependency, with greater statistical difference between conditions (i.e., pre and post arousal) for the elderly. Auto-mutual information revealed shorter decay times for the periods following arousal, indicating diminished predictability in cortical dynamics, i.e., more irregularity. Remarkably, the previous differences were not detectable through power spectral changes, and were only detectable using the mutual information indices. The authors speculate at the neurophysiological level that lower interdependency, as reflected by reduced cross-mutual information, could render the brain more vigilant and vulnerable to arousal disturbing issues affecting sleep continuity in the elderly.

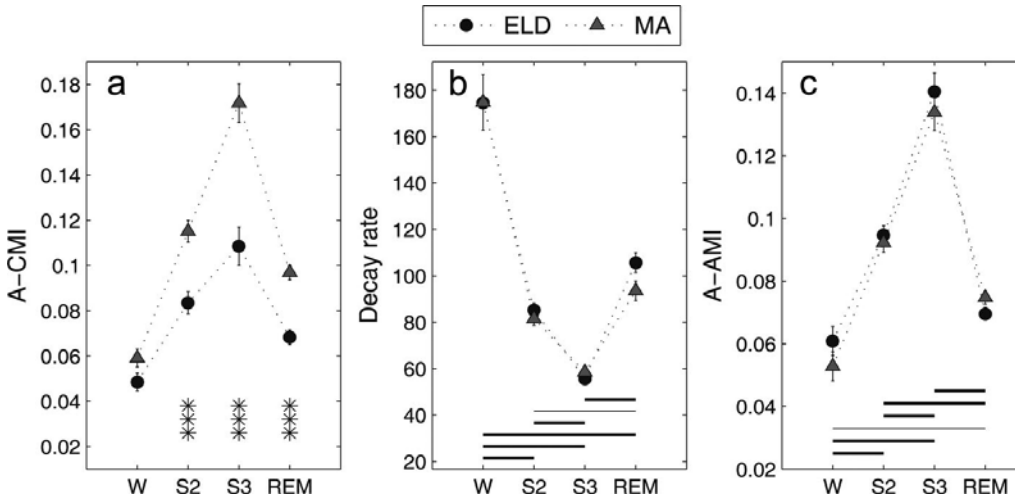


***Fig. 5.3.*** *Information theory based measures to study differences in sleep patterns between middle-age (MA) and elderly (ELD) women. Left: mean values and standard error of the group average cross-mutual information (A-CMI); center: decay rate of the AMI; and right: average auto-mutual information (A-AMI). See text for a complete description of the statistical tests.* Reproduced from (Ramanand et al., 2010a).

## 5.3. Effective connectivity (EC)

### 5.3.1. Transfer entropy (TE)

Transfer entropy (derived from information theory; Table 5.1) is a non-linear measure, asymmetric in its variables (hence, able to discriminate the dominant directionality), that aims to quantify the causal transfer of information between two signals. A more rigorous definition is that transfer entropy from a signal *X* to a signal *Y*, represents the reduction of uncertainty in the prediction of the current value of *Y*, first using only past values of *Y*, and afterwards including past values of *X* as well (Vicente et al., 2011, Lee et al., 2012, Huang et al., 2015).

However, despite a technical superiority and suitability to investigate all kinds of causal interactions (including the highly non-linear ones), the theoretical basis might become mathematically challenging, with the additional caveat of being computationally demanding (Barrett et al., 2012, Silfverhuth et al., 2012, Chan et al., 2013). Given the fact that easier linear methods (e.g., Granger causality or Partial Directed Coherence) are also available to tackle the problem of effective connectivity, some authors consider them as preferable over transfer entropy or other variants (such as Renyi's information transfer) (Astolfi et al., 2007, Silfverhuth et al., 2012). Nonetheless, it can be demonstrated that, in case of Gaussian distributed variables, transfer entropy and Granger causality are equivalent methods (Barnett et al., 2009, Seth et al., 2011). Reviewing neuroscientific literature on this topic, most applications using information theory pertain to functional connectivity and only a minority to effective connectivity. Nevertheless, the latter is a field that has gained greater interest in recent years (Vicente et al., 2011, Lee et al., 2012, Wibral et al., 2014, Wollstadt et al., 2014, Huang et al., 2015).

### 5.3.2. Granger causality (GC)

Granger causality (GC) is a statistical technique that was firstly introduced by Nobel prize winner in economics Sir Clive Granger in 1969 (Granger, 1969), initially applied in his studies in the field of econometrics. In essence, GC is a technique based on autoregressive linear modeling for detecting effective connectivity, and hence detecting the direction of propagation of information flow between brain regions (Ashley and Tsang, 2014, Bastos and Schoffelen, 2015). GC has been applied in recent years to include many more scientific areas than economics with similar success, particularly including computational neuroscience (Sameshima et al., 1996, Sameshima et al., 1998, Bernasconi and König, 1999, Kaminski et al., 2001).

In practice, GC arises naturally in computational neuroscience when we encounter the problem to best predict the future of a particular EEG channel (for instance, by means of a Multivariate Autoregressive (MVAR) modeling; (Erla et al., 2009, Faes et al., 2010)), by using either the entire rest of channels, including past values of our chosen one, or after discarding certain number of targeted channels (those in which we desire to test a causal relationship) (Faes and Nollo, 2013). The appropriate order of a general MVAR model can be determined following the Akaike Information Criteria, (Akaike, 1974). It should be mentioned that, in general, minor changes of the MVAR order do not exert a significant impact on the final results (Korzeniewska et al., 2003, Astolfi et al., 2007). If the predictions are worse after excluding the group of targeted channels, then it is said that these excluded channels have established a “G-causal” relationship with the originally selected channel (Barrett et al., 2010). The previous definition applies generally, but standard analyses using GC generally proceed in a pair-wise fashion, that is, focusing on a single targeted channel (which can vary) with respect to its influence on the selected main channel (i.e., two variable models) (Sommerlade et al., 2012).

Perhaps the best possible definition the one of Granger himself (Granger, 1969):

*The basic "Granger Causality" definition is quite simple. Suppose that we have three terms, $X_t$, $Y_t$, and $W_t$, and that we first attempt to forecast $X_{t+1}$ using past terms of $X_t$ and $W_t$. We then try to forecast $X_{t+1}$ using past terms of $X_t$, $Y_t$, and $W_t$. If the second forecast is found to be more successful, according to standard cost functions, then the past of $Y_t$ appears to contain information helping in forecasting $X_{t+1}$ that is not in past $X_t$ or $W_t$. In particular, $W_t$ could be a vector of possible explanatory variables. Thus, $Y_t$ would "Granger cause" $X_{t+1}$ if (a) $Y_t$ occurs before $X_{t+1}$; and (b) it contains information useful in forecasting $X_{t+1}$ that is not found in a group of other appropriate variables.*

*Naturally, the larger $W_t$ is, and the more carefully its contents are selected, the more stringent a criterion $Y_t$ is passing. Eventually, $Y_t$ might seem to contain unique information about $X_{t+1}$ that is not found in other variables which is why the "causality" label is perhaps appropriate.*

Hence, with GC, one can state that one time series $Y_t$ helps in the forecasting of future values of another time series $X_t$. In this case, signal $Y_t$ “Granger causes” the dynamics of signal $X_t$, and even if its variations are heavily correlated, rigorously we cannot claim $Y_t$ is actually causing $X_t$ evolution for sure (Baccala and Sameshima, 1999, Faes et al., 2010). For instance, we might be incurring in an omitted-variable bias and obtaining a spurious relationship, ignoring hidden variables (so called “confounding variables”, relating to the vector referred as $W_t$) that bridge the causation between $Y_t$ and $X_t$ (perhaps, even with no interaction at all between them) (Kaminski et al., 2001, Friston et al., 2014). This is something occurring not only in economic applications, but also in the context of neuroscience (e.g., regarding the activation among distinct regions of the cortex) (Sameshima et al., 1998, Baccala and Sameshima, 2001). Consequently, the fact that two brain regions get activated temporally proximal, even systematically, is not implying a direct communication between them; there could be a

third region -the confounding variable- that is actually activating them with the introduction of some temporal lag (Kaminski et al., 2001, Faes and Nollo, 2013).

In the current network-centric paradigm of effective connectivity (as opposed to the traditional paradigm of strict functional localism in the brain (Roy, 2012)), conceptualized as a stochastic, non-deterministic, set of connected nodes, the same underlying network structures of the cortex are thought to be able of generating different causal interconnections, depending on the context (i.e., external environment as well as other inner factors) or the type of cognitive task (Astolfi et al., 2007). As a consequence, GC is receiving a lot of attention among researchers as an easy-to-apply, yet efficient technique allowing a satisfactory dynamical description (as it evolves with time) of information flow in stochastic and transient networks formed by the brain (where its emergent responses may vary even if the presented stimulus is the same) (Sommerlade et al., 2012, Friston et al., 2014). For this main reason, GC has been applied to many neuroscientific problems in recent years (Ashley and Tsang, 2014, Faes et al., 2014, Seth et al., 2015), showing up as a very fruitful and efficacious technique for the identification of “causal interactions” underlying functional connectivity (Seth et al., 2015).

### 5.3.3. Applications to loss of consciousness (LOC)

Barrett and colleagues applied GC to investigate the brain changes in a state of loss of consciousness (LOC) induced by the anesthetic Propofol, in comparison to wakeful resting (WR) (Barrett et al., 2012). Though this altered state of consciousness is not exactly sleep, it shares many features in common with slow wave sleep (Mashour, 2011, Murphy et al., 2011a). The study was performed in 7 healthy subjects. rEEG sources of two particular regions of interest (ROIs), the anterior cingulate cortex (ACC) and the posterior cingulate cortex (PCC) were determined. These brain areas were chosen as they are implicated in the mesial pathway responsible for slow wave propagation, both in anesthesia and sleep.

GC analyses in different frequency bands delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (16-25 Hz) and gamma (25-40 Hz) were performed. The time domain GC was also calculated, as an average integral from 0 to the Nyquist frequency of the GC spectral representation. GC from ACC to PCC and vice versa were determined.

GC (PCC to ACC) in delta band exhibited a significant increase in anesthesia compared to wake (Fig. 5.4a; false discovery rate lower than 5%) in subjects 2, 4 and 6, in the theta band in subjects 1, 2 and 5, in the alpha and beta band only in 2 and 5, and in the gamma band in 1, 3 and 7 (exceptionally, subject 3 showed a decrease). The GC in the time domain showed a significant increase in subjects 1, 2 and 5.

GC (ACC to PCC) in the delta band did not show significant differences between anesthesia and wake, in the theta and alpha band GC was higher in anesthesia only in subject 4, in the alpha and beta band in 2, 4, 6 and 7, and in the gamma band in all subjects except 3. The GC in the time domain showed a significant increase in subjects 2, 5 and 6.

Overall, GC reveals that the main effect of LOC induced by anesthesia in most subjects consists in significant increases of bidirectional causal flow connecting the ACC and PCC, being especially prominent in the high frequency range (beta and gamma bands).

To conclude, GC is an excellent method to be included in sleep studies (both for human and animal studies; some applications to rats can be found in (Sameshima et al., 1996, Sameshima et al., 1997)), as it detects causal effective connectivity between signals very effectively, being much simpler both conceptually and computationally than other non-linear methods (Astolfi et al., 2007, Liu and Aviyente, 2012), having the additional advantage of requiring less data (Ashley and Tsang, 2014). Lastly, as an extension GC effective connectivity can be analyzed as a function of frequency (spectral G-causality, where the associated MVAR model is transformed to the Fourier space, using the transfer function matrix, (Faes and Nollo, 2013)), an advantage in the context of sleep.

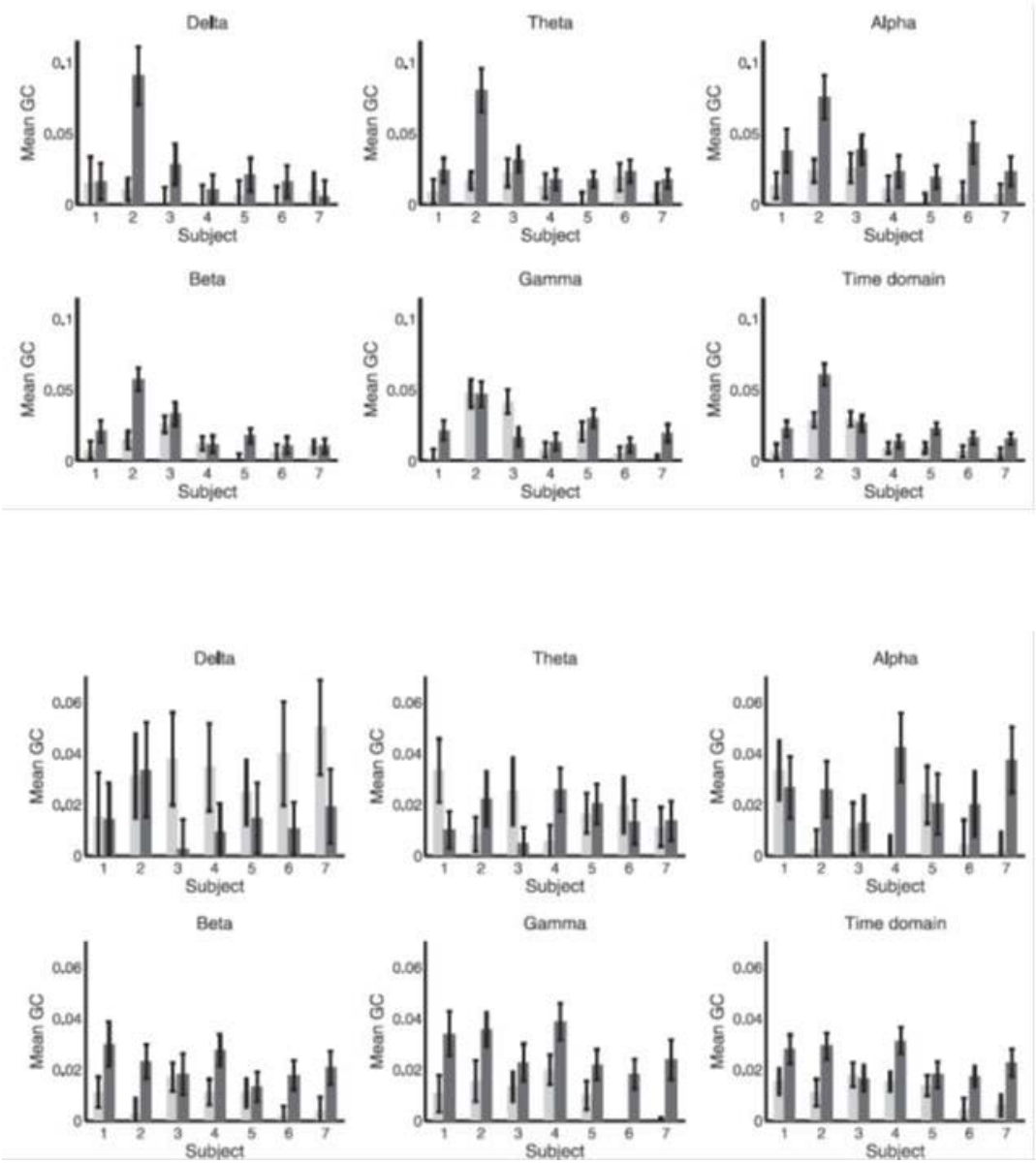


***Fig 5.4. a. (upper)*** *GC (mean values and standard errors) from PCC to ACC and* ***b. (lower)*** *from ACC to PCC in all subjects and different frequency bands, as well as its time domain version. Grey (left bar) corresponds to waking condition, while dark (right bar) to LOC.* Reproduced from (Barrett et al., 2012).

### 5.3.4. Directed Transfer Function (DTF), Partial Directed Coherence (PDC) and Isolated Effective Coherence (iCOH)

What has been discussed so far to study directionality is well suited for the comparison of influences between pairs of channels. Nonetheless, one may ask whether methods able to deal with higher dimensionalities (i.e., multiple channels simultaneously) also exist. Several statistical techniques have been proposed that have gained popularity to tackle the problem of multichannel effective connectivity (Erla et al., 2009, Faes et al., 2012, Sommerlade et al., 2012, Karlsson et al., 2013, Khadem and Hossein-Zadeh, 2014). However, here we will concentrate on three important methods: the Directed Transfer Function (DTF, (Kaminski and Blinowska, 1991)), the Partial Directed Coherence (PDC, (Baccala and Sameshima, 1999), (Baccala and Sameshima, 2001)) and the Isolated Effective Coherence (iCOH, (Pascual-Marqui et al., 2014a)), as being highly efficient and useful in computational neuroscience.

All these measures have in common that they stem from GC theory (hence, they constitute linear estimators of connectivity), but extending and generalizing it by using MVAR models in order to include multichannel data (Bernasconi and König, 1999, Blinowska, 2011, Silfverhuth et al., 2012). In principle, the associated theory poses no restriction on the number of channels that can be included, but in practice, the corresponding computational load (particularly with high-density EEG) puts some practical limits, depending on the available computational power (particularly demanding in the case of the DTF, by requiring a matrix inversion whose order is equal to the number of channels (Baccala and Sameshima, 1999)).

A second commonality of these methods is that they are mathematically formulated in the Fourier space, therefore, representing effective connectivity values as a function of frequency, a feature particularly interesting for sleep studies (Kaminski et al., 1997, Faes et al., 2012, Pascual-Marqui et al., 2014a). In contrast to functional connectivity methods in the frequency domain (such as coherence), it is worthy to remember that methods addressing effective connectivity

break the symmetry of an interacting pair (i.e., in a matrix representation, the permutation of indexes will change the values). Biologically, this non-symmetric character can be interpreted as a hierarchical relationship even for a single pair of channels, with one of them being more dominant in its causal influence over the other one (Schelter et al., 2006, Stam et al., 2016).

Finally, the methods also share the normalization, with the measures ideally ranging from 0 (i.e., no causal relationship between a given pair of channels in the network) to a maximum of 1. This range has the advantage of being straightforwardly interpreted as percentages of causal influences (Baccala and Sameshima, 2001, Bastos and Schoffelen, 2015). However, since the presence of noise and the finite nature of the signals will always render values larger than strictly zero. Thus, statistical comparison criteria are unavoidable (e.g., establishing a bootstrap distribution to obtain a threshold of significance), in order to discard non-significant interactions from the true causal influences (Astolfi et al., 2007, Faes et al., 2009, Blinowska, 2011).

For an effective connectivity representation including both frequency and time, an analysis of consecutive epochs (short time segments, typically of a few seconds), comparable to the approach followed by the short time Fourier Transform in a classical spectrogram, would be an approach, as e.g. implemented in the short-time DTF (Philiastides and Sajda, 2006, Korzeniewska et al., 2008, Blinowska, 2011). The longer the epoch, the more accurate the estimation of connectivity indexes (additionally, the MVAR fitting would become more precise, at least if the epoch length is still short enough to assume quasi-stationarity, with approximately stable mean and variance), but worsening temporal resolution, and vice versa (Barrett et al., 2012). This behavior is in correspondence with the well-known trade-off between time and frequency resolution in signal processing (Roach and Mathalon, 2008).

In the special case of event-related potential (ERP) experiments with multiple trials, the possibility to perform an average over repeated trials has the remarkable advantage of improving MVAR based

effective connectivity estimators without compromising the temporal resolution, thus, attaining greater statistical power (Liang et al., 2000, Philiastides and Sajda, 2006, Blinowska, 2011). As a final consideration, the number of samples in an epoch should be at least a few times greater (depending linearly - on the model order used for the MVAR) than the total number of channels included (Korzeniewska et al., 2008, Karlsson et al., 2013).

Sometimes, the DTF and the PDC (see below) are seen as complementary methods, since the flaws of one technique are not present in the other, and vice versa (Astolfi et al., 2007, Silfverhuth et al., 2012). Moreover, the mathematical formulation and interpretation of these indexes can be understood as being reciprocal to each other (Faes et al., 2012). In simple terms, the DTF quantifies the amount of causal influence between a source channel to a particular sink channel, normalized with respect of all causal inflow the sink is getting from the entire network (including itself) (Kaminski et al., 2001, Babiloni et al., 2005). Thus, the DTF can be used to find which ROI among a set of candidates is the most responsible one for the causal interaction with a given brain area (the sink channel), by maximizing this variable. Reciprocally, the PDC measures the amount of causal influence between a source channel and a given sink channel, normalized with respect to the total outflow the source is transmitting to the entire network (including itself) (Baccala and Sameshima, 2001, Huang et al., 2016). Mathematically, a characteristic they share is that the squared sum of elements along a row or a column (for the DTF and the PDC matrices, respectively) is always equal to 1, a property that can be used to achieve a modest reduction in computational load for network analyses (Blinowska, 2011).

Additionally, the DTF, being a straight forward normalized generalization of GC, still combines direct influences with indirect cascade pathways linking two arbitrary channels in the network (Djordjevic et al., 2012). This is a consequence inherent to the mathematical definition of the DTF, which involves a matrix inversion for the transfer function of the system (Kaminski et al., 2001, Astolfi et

al., 2007, Korzeniewska et al., 2008). Conversely, since the PDC does not require any matrix inversion, it yields only the direct causal connections (hence, having better specificity than the DTF), and at the same time, significantly reduces computational load for not requiring this demanding step of a matrix inversion (Baccala and Sameshima, 2001, Schelter et al., 2006, Schelter et al., 2009). However, this property has to be counterbalanced with the fact that the interpretation provided by the DTF is in many contexts preferable to the PDC, since it is explicitly designed to answer the question of how one particular ROI is causally affected by the rest of the network (Korzeniewska et al., 2003).

On the other hand, the iCOH tries to overcome limitations present in both previous methods, but its mathematical definition is much closer to the PDC (although requires more theoretical considerations to define it) than the DTF, also involving no matrix inversion (Pascual-Marqui et al., 2014a, Huang et al., 2016). Basically, the iCOH changes the normalization of the PDC, leaving invariant the numerator (that is, the connection between a source and a sink channels) but eliminating from the denominator all connections representing transmission from the source to a different sink, except for the particular channels involved (thus, only remaining two terms) (Pascual-Marqui et al., 2014a, Cao et al., 2016). Additionally, weighting factors (inversely proportional to the error from the MVAR equations fitting) are included to achieve better accuracy. The main motivation for introducing the previous change in normalization is manifested in the property of iCOH not lowering progressively its values (possibly, even getting very close to zero) as it happens with the PDC and the DTF if additional new channels are included in the analysis (remaining the former channels unaltered) (Pascual-Marqui et al., 2014a, Lin et al., 2015).

A brief but important question concerning applications of these methods to EEG data refers to the issue of volume conduction effects (Pascual-Marqui et al., 2011, Bastos and Schoffelen, 2015). Fortunately, it has been demonstrated that this problem can be simply ignored in the context of our discussion, as volume conduction does not affect the final indexes results by respecting the relative phase

relationships in the network (De Gennaro et al., 2004, Kaminski and Blinowska, 2014). Given that volume conduction does not alter the phase difference between derivations (i.e., not that individual phase signals are not affected, but that they keep invariant its relative difference), preprocessing in this respect (commonly achieved, e.g., by means of the Laplace transformation or Independent Component Analysis) should be avoided, as some part of the original structure of the data is systematically modified (possibly even destroyed) by these preprocessing techniques (Kaminski and Blinowska, 2014). The only preprocessing technique that could be helpful and recommended in this regard (prior to the application of a MVAR model, on which these methods are built) is the z-score normalization of all channels, that is, subtracting the mean value of each channel and dividing by the corresponding standard deviation (Kaminski et al., 2001, De Gennaro et al., 2004).

It must be stressed that the reason for presenting techniques able to work with higher dimensionalities (that is, including multiple channels simultaneously) is not just a matter of a mere optional mathematical refinement. There are many computational and, above all, theoretical advantages to support this sort of extension, where MVAR based models acquire full flexibility and applicability (Erla et al., 2009, Faes et al., 2012, Karlsson et al., 2013). As a first illustrative example, consider the network referred below in Fig. 5.5. Using pairwise analysis of GC, it can be mathematically proven that this algorithm, in its original formulation based upon a bivariate AR model, does not distinguish properly between these two possible connectivity scenarios.

In effect, the classical GC method cannot disentangle correctly between these very simple yet different network configurations, as a connectivity pattern based on $X \rightarrow Z \rightarrow Y$ yields as a result a non-zero value of GC (i.e., a false positive) between extreme channels $X$ and $Y$, when this causal interaction may or not actually exist (Kaminski et al., 2001, Faes et al., 2010). In fact, this is a general pitfall occurring when using bivariate analyses of functional or effective connectivity, that is, when including in the model just a pair of channels, and hence ignoring

the structural information contained in the rest of the network (Faes et al., 2012, Sommerlade et al., 2012). More generally, this limitation has also been demonstrated for a straight generalization of GC to multiple dimensions (suitable for multichannel data), where both false positives (type 1 error) and false negatives (type 2 error) in cascade flows are due to the lack of proper distinction between direct and indirect causal pathways connecting a given pair of channels (Kaminski et al., 2001, Blinowska, 2011).

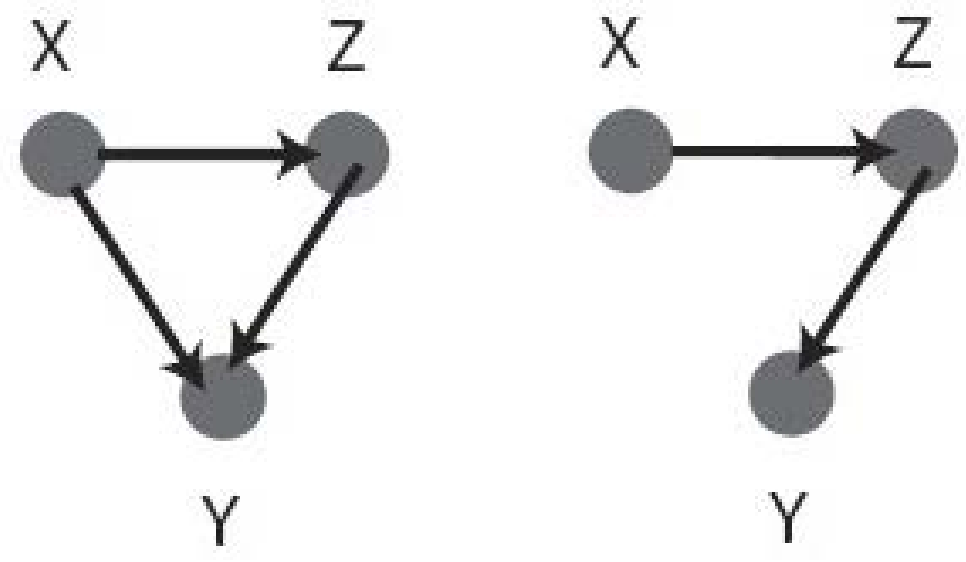


***Fig. 5.5.*** *A simple network to demonstrate the limitations of bivariate analyses of effective connectivity in neuronal networks.*

Thus, the previous example serves to illustrate the problematic ambiguity resulting from analyses conducted only in a two-dimensional space, thus, showing how cautious one must be with these analyses seeking for the actual and direct effective connectivity (Pascual-Marqui et al., 2014b). In contrast, full multivariate based analyses (as provided by DTF, PDC and iCOH) are significantly better in its accuracy at differentiating between these possible scenarios (nonetheless, no connectivity method is completely flawless), both in terms of sensitivity and specificity for the genuine network connections (Astolfi et al., 2007, Silfverhuth et al., 2012). Consequently, the addition of complexity provided by these multivariate methods compared to the classical GC pays off based on the richness of the benefits they reap, since they naturally extend GC power to whole networks (Philiastides and Sajda, 2006, Liu and Aviyente, 2012). Simultaneously, in the process, they overcome many of the drawbacks present in only two-dimensional analyses, where mathematical simplification implies not only ignoring valuable information, but directly getting wrong conclusions in many cases; under this dichotomy, simplifications are not recommended (Kaminski et al., 2001, Blinowska, 2011).

As a second example of possible failures associated with bivariate analyses, GC would falsely conclude causality between two signals if they are actually being activated or controlled by a third variable, but are time delayed, e.g. due to differences in the distances to travel (a common case for brain networks, thus yielding spurious results at causal detections) (Kaminski et al., 2001, Friston et al., 2014). Therefore, bivariate GC would proclaim effective connectivity from the signal (or brain structure) with the shorter temporal delay towards the one with a larger delay (assuming causality propagates forward in time, by definition), thus producing a false positive causal interaction, even when there are no indirect paths of causal linkage (Granger, 1969, Kaminski et al., 2001).

In summary, the main advantage of these MVAR based methods is their ability to efficiently detect the directional influences and

concomitant interaction strength between every possible combination of pairs of variables in a multivariate dataset, theoretically, including all recorded channels for analysis simultaneously (ignoring the issue concerning the resultant computational load) (Karlsson et al., 2013). As explained before, this extension poses a double benefit: firstly, avoiding a daunting, pairwise sweeping set of analyses that grows geometrically with the number of channels $N$ (approximately as $N^2$) and, secondly, but more importantly, circumventing the theoretical flaws arising in bivariate AR based models, when available information from the network channels, other than the current pair, is suppressed in the calculations (Blinowska, 2011, Faes et al., 2012, Khadem and Hossein-Zadeh, 2014).

Despite of having common roots inside their mathematical structure, in a more detailed comparison, some noteworthy differences start to emerge. For instance, regarding the tolerance to low signal-to-noise-ratio (SNR), the DTF exhibits the best robustness against noise and the consequent corruption of the EEG signals, but this superiority does not surpass to a great extent that of the other methods (Astolfi et al., 2007, Silfverhuth et al., 2012). The DTF also has the advantage of good consistency, that is, it remains basically invariable regardless the EEG electrode reference employed for analysis (De Gennaro et al., 2004, Astolfi et al., 2007, Blinowska, 2011). Nonetheless, the average reference is generally not recommended to be used in MVAR based connectivity methods, since the latter operation mixes signals from all channels, therefore, potentially yielding to some spurious connections (Kaminski et al., 2001, Astolfi et al., 2007).

Another relevant feature to be considered is the minimum necessary data length to get reliable estimates (assuming ideal stationary conditions) (Karlsson et al., 2013). Using synthetic networks with predefined connectivity parameters, Astolfi et al. (2007) showed that a minimum number of about 6,750 samples (corresponding to 27 s of data sampled at 250 Hz) led to an error below 7% in detection of causal connections for the discussed methods, and less than 5% in the case of the DTF (Astolfi et al., 2007). However, in spite of the superiority of the DTF with respect to data length and SNR, the PDC,

and particularly the iCOH, perform generally better at detecting direct connections and respecting the original frequency transmission responses (Silfverhuth et al., 2012, Pascual-Marqui et al., 2014a, Huang et al., 2016). Thus, the DTF is more likely to render both indirect connectivity paths along with the true paths, and also to miss some direct connections (Djordjevic et al., 2012).

Even with its inferiority in terms of sensitivity and specificity regarding causal connections when compared to the other methods, the DTF is preferable than the use of GC, even when introducing a multichannel generalization (non-normalized) of GC (Kaminski et al., 2001, Philiastides and Sajda, 2006). This is a consequence of two facts: firstly, for the standard GC, the theory introduces a bivariate autoregressive model, but the DTF uses a full MVAR of all channels, automatically taking into account the whole structure of the network (Kaminski and Blinowska, 1991, Blinowska, 2011). Secondly, even when generalizing GC to multiple channels in the frequency domain (similar to a non-normalized DTF), the normalization factor present in the denominator of the DTF helps to attenuate the number of false positives that GC would predict, hence improving statistical analyses for true causality (Blinowska, 2011). For a final analysis of statistical significance, it was recommended to threshold these connectivity indexes (since conveniently all of them are normalized between 0 and 1) around 0.2 (Djordjevic et al., 2012), or alternatively to use a bootstrap analysis with surrogates of the signals (e.g., 500 surrogates), and then selecting the top 1-5% of strongest connections of this empirical distribution (Astolfi et al., 2007, Faes et al., 2009).

Due to the flaws of DTF, but knowing of its superiority with respect to data length and SNR, a modification of the technique to improve its sensitivity and specificity has been proposed, called the directed Direct Transfer Function (dDTF) (Korzeniewska et al., 2003). The improvements consist firstly, of the integration of the denominator of the DTF over all frequencies, to get the so called full-frequency DTF (ffDTF), and secondly, by multiplying the latter parameter with the partial coherence (not to be confound with the PDC) (Blinowska, 2011). Hence, the information extracted with the dDTF is superior to

the DTF, as it is merged with results derived from a functional connectivity technique (the partial coherence), in order to better differentiate the true causal connections (Korzeniewska et al., 2003).

The partial coherence is a frequency dependent measure normalized between 0 and 1 removing or discounting the influence effects of the rest of the network (Astolfi et al., 2007, Pascual-Marqui et al., 2014a). Thus, partial coherence, and not the classical coherence, is the right measure to use in the context of multichannel data and network structure (Baccala and Sameshima, 1999, Korzeniewska et al., 2003).

Therefore, it is after combining the ffDTF with the partial coherence that the dDTF is able to better reveal causal relationships underlying the network structure, widely diminishing the main limitations of the DTF regarding false positive causal relationships (Korzeniewska et al., 2008). However, a caveat of the dDTF it that due to its inclusion of the partial coherence (as previously said, normalized between 0 and 1), but mainly for the frequency-integrated denominator (appearing in the ffDTF), the dDTF may sometimes render too low values, unless the causality link is particularly strong (Astolfi et al., 2007). DTF would be in this sense more exact, so neglecting the DTF in favor of the dDTF would eliminate the false positive, but bias the relative strength of the connections to appear weaker than they actually are; for such ambiguous cases, comparison with the original DTF would be strongly recommended (Astolfi et al., 2007).

It must be emphasized that all the methods presented in this section complement each other to gain a more complete understanding of problems arising in computational neuroscience. It is not about competing with each other just for the sake of superiority (as authors sometimes do).

### 5.3.5. Applications to sleep studies and related areas

An advantage of sleep studies that makes them privileged in comparison to wake EEG studies is that sufficiently long and reasonably stationary data series are available for analyses. Consequently, the application of MVAR based methods becomes quite straightforward (Kaminski et al., 2001, Karlsson et al., 2013).

Kaminski et al. (Kaminski et al., 1997) applied the DTF to investigate effective connectivity patterns aiming to differentiate between sleep stages and wakefulness. In this study, whole night recordings with 21 electrodes (10-20 international system, linked ears reference) in eight healthy participants were performed. The DTF was calculated in the range of 0-30 Hz for all pairs of channels (10-s epochs averaged across stages). Since with an array of 21 electrodes, there are 210 possible pairs of connections (considering also directionality), a threshold needs to be taken into account in order to discount non-significant connections, which was set at 0.35 (based on a typical histogram; the DTF is normalized between 0 and 1 by definition).

The DTF in the range of 0-30 Hz showed propagation of activity from fronto-central regions towards the rest of cortical structures during deep sleep (Fig. 5.6). However, this study was based on scalp analyses, and thus, although tantalizing, conclusions should not be confused with rigorous results provided by connectivity analyses on inverse solution of the EEG (Pascual-Marqui et al., 1994). The causal range of influence was highest during waking followed by deep sleep (stages 3 and 4).

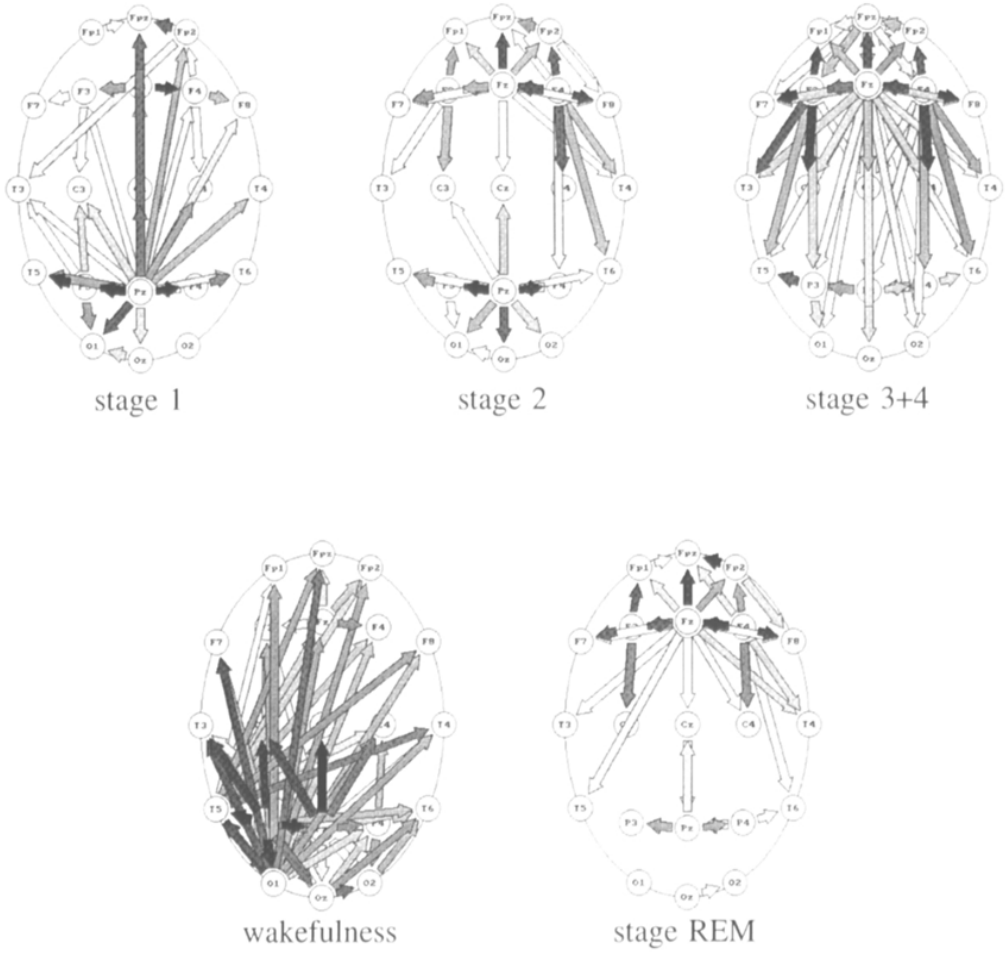


***Fig. 5.6.*** *Example of the DTF method applied to different sleep stages and wakefulness, integrated over 0 and 30 Hz. Darker arrows reflect higher DTF values.* Reproduced from (Kaminski et al., 1997).

Topographical networks provided by the DTF were consistent among subjects across stages, exhibiting only minor inter-individual differences (for example, posterior centers of propagation could be shifted toward occipital derivations in some individuals, but toward parietal electrodes in others). The highest complexity appeared during waking, with principal sources of activity mainly located in posterior regions. During stage 1, the major source hub appeared around Pz. On the other hand, during stage 2, sources of activity were frontal (F3 and F4, Fz) and posterior (Pz), furthermore, displaying high interhemispheric symmetry. During deep sleep, frontal sources prevailed, while their range of causal propagation spanned basically the entire cortex. During REM sleep, the range of interaction of frontal sources decreased compared to other sleep stages.

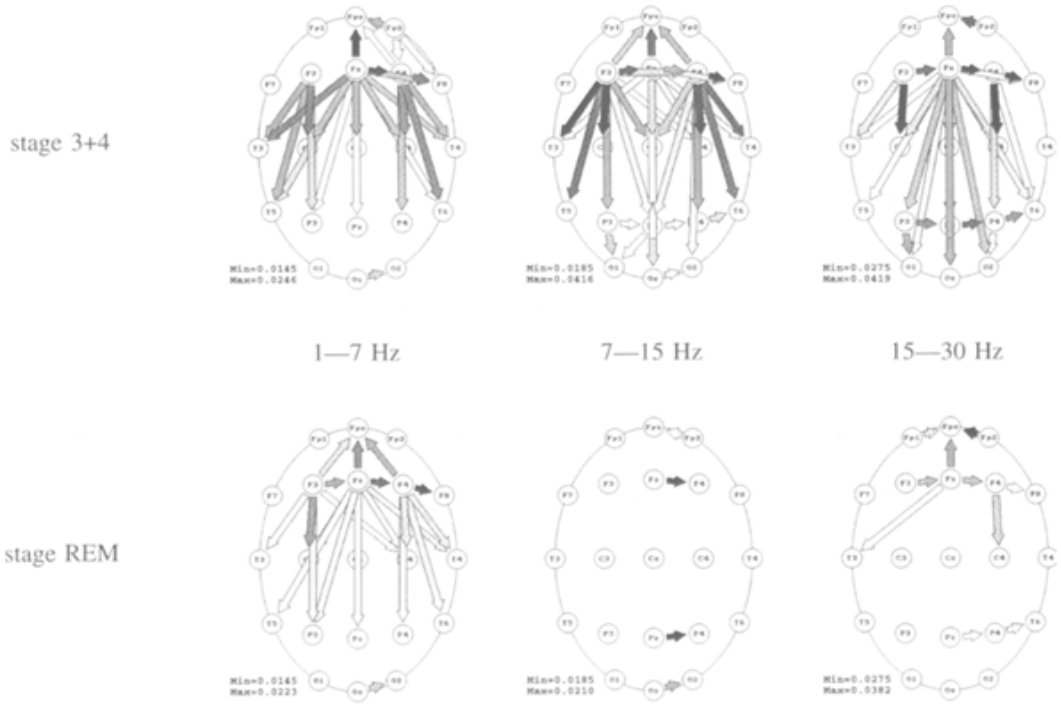


***Fig. 5.7.*** *Example of the DTF method applied to stages 3 and 4 and REM sleep, integrated over 1-7, 7-15 and 15-30 Hz. Darker arrows reflect higher DTF values.* Reproduced from (Kaminski et al., 1997).

DTF topographical maps integrated in three different frequency bands (1-7 Hz, 7-15 Hz and 15-30 Hz) are illustrated in Fig. 5.7. In deep sleep (Fig. 5.7, upper row), major sources were in frontal areas for all frequency bands with rather symmetrical connections.

In REM sleep (Fig. 5.7, lower row), connections were basically restricted to the low frequency band with a frontal source.

In a later study, Kaminski and colleagues (Kaminski et al., 2001) calculated the connectivity patterns of the human cortex during sleep stage 2. Application of the DTF (7-15 Hz) revealed primarily frontal sources and more connections in the left hemisphere (Fig. 5.8). No interhemispheric connections were observed.

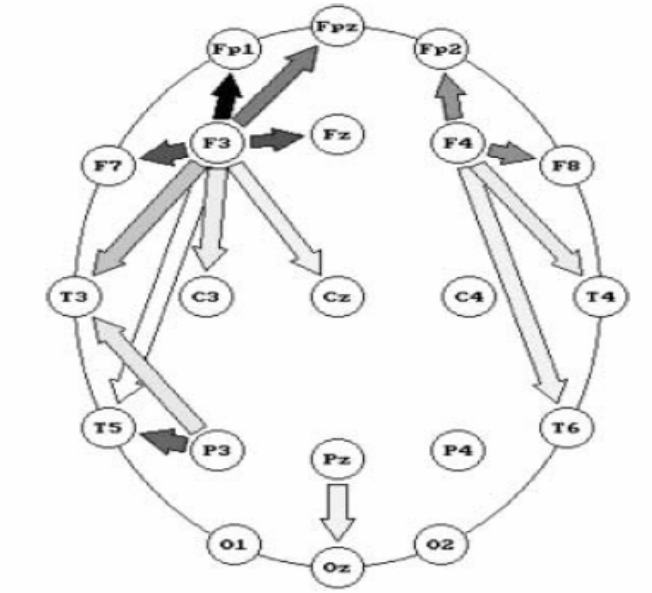


***Fig. 5.8****. Example of the DTF method applied to stage 2, integrated from 7 to 15 Hz. Darker arrows reflect higher DTF values.* Reproduced from (Kaminski et al., 2001).

Furthermore, Kaminski et al. (2001) reported increasing anterior to posterior causal influences during the transition from waking to sleep. The main causal sources were located around channels F3 and F4 for anterior regions (indicating prefrontal cortex projection pathways), and P3 for the left posterior region. The causal influences were exerted mainly within each hemisphere, displaying short and medium range connectivity network patterns.

Continuing with applications of the DTF to sleep, De Gennaro and colleagues studied the sleep onset transition in humans by means of this technique (De Gennaro et al., 2004). EEG activity of 3 derivations covering a fronto-posterior network Fz, Pz and Oz (linked-mastoid reference) were analyzed (Fig. 5.9). The focus was on the investigation of the causal influences in the fronto-parietal-occipital network at the sleep onset transition, defined as the first emergence of stage 2 epoch (time prior to this moment is referred as “presleep” condition). It was hypothesized that dorsolateral prefrontal areas may play a leading role in the transmission of synchronizing signals to the occipital cortex, explaining hypnagogic imagery hallucinations accompanying the sleep onset process (De Gennaro et al., 2004, Ong et al., 2015).

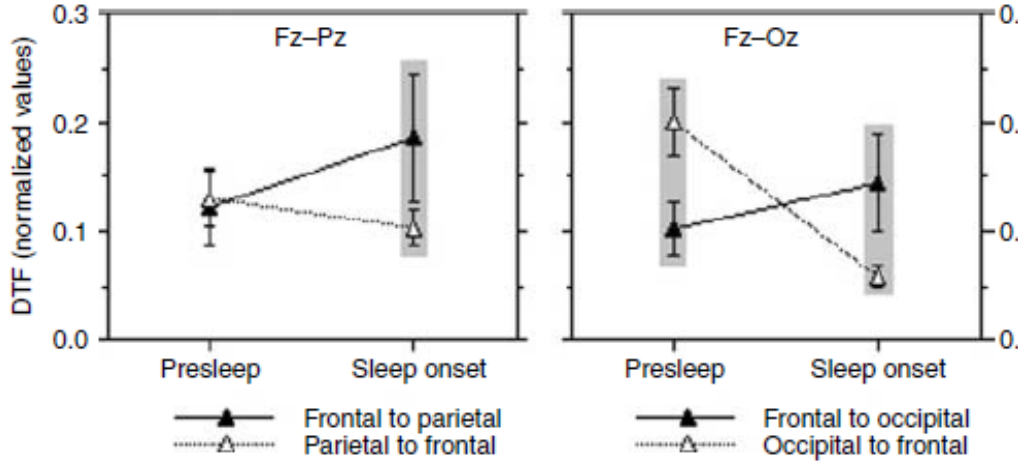


***Fig. 5.9.*** *Mean DTF values across subjects (10 participants) and standard errors (vertical bars), indicating the conditions (presleep and sleep onset) and the directionality for the pair of electrodes selected (from frontal to parietal and its reciprocal in the left panel, Fz-Pz, and from frontal to occipital in the right panel, Fz-Oz). The DTF values were averaged from 1 to 28 Hz. Dashed areas represent statistical significance according to Duncan post hoc comparison (from left to right, P=0.0007, 0.0003 and 0.0006).* Reproduced from (De Gennaro et al., 2004).

A statistical comparison revealed that coherence in the low frequency range (delta and theta bands) was higher in the period preceding the sleep onset transition (presleep) than after it, however, the opposite behavior was true for the alpha band, with higher coherence after sleep onset. The DTF showed a prevalence of occipital to frontal information flow in the delta, theta and alpha bands for the presleep condition. On the other hand, after sleep onset, the DTF results indicated a pattern of frontal to parieto-occipital flow significant for all frequency bands, suggesting a prominent role of prefrontal regions in the propagation of synchronizing signals spanning delta to beta bands.

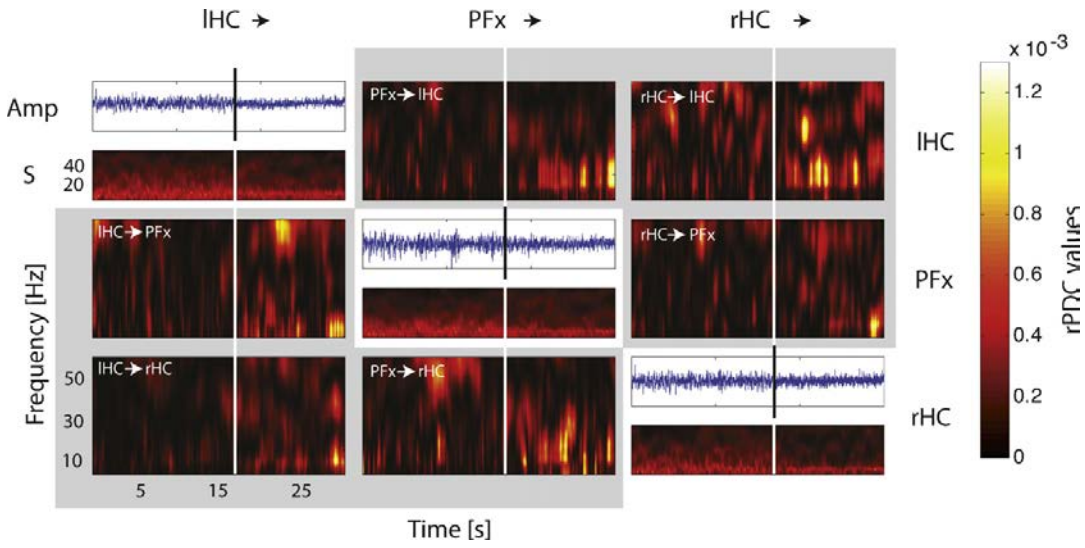


***Fig. 5.10.*** *Application of the renormalized PDC (rPDC) to the slow wave sleep to REM sleep transition in mice. Vertical lines indicate the transition (at 16.7 s). Diagonal elements represent the EEG (top) and the corresponding spectrograms (bottom); off-diagonal elements correspond to the rPDC graphs as a function of time and frequency for pairs of regions (left and right hippocampus and prefrontal cortex).* Reproduced from (Sommerlade et al., 2012).

Sommerlade and colleagues (Sommerlade et al., 2012) analyzed the transition from slow wave sleep to the REM sleep using the PDC technique (more precisely, it is the renormalized PDC, or rPDC, proposed in (Schelter et al., 2009)). Activity was recorded from the medial prefrontal cortex (PFx), left and right hippocampus (lHC and rHC).

The Fig. 5.10 summarizes the rPDC results. The lines represent the transition point from slow wave to REM sleep, occurring at 16.7 s. Before the transition, hardly any interactions were revealed, and if present, they were very transient and irregular. In contrast, after the transition from slow wave to REM sleep, the rPDC plots show significant interactions between the considered brain regions. The strongest transmission occurred from rHC to lHC. Initially after the transition, a causal interaction can be observed in the gamma band (centered around 40 Hz), then followed by a more stable and longer interaction that ranges from 10 to 30 Hz (upper beta). Towards the end of the recording, the direction was inverted (lHC to rHC, alpha band),

as it also happened for the interactions from PFx to lHC, and rHC to PFx. The earliest influence showing up appeared in the gamma band (frequencies centered around 50 Hz) in the link that connects lHC to PFx, lasting a few seconds.

Finally, an application of iCOH to human EEG data was conducted by Pascual-Marqui and colleagues (Pascual-Marqui et al., 2014a). As there are currently no publications of this method in the sleep field we summarize the iCOH application to EEG recordings with eyes open (EO) and eyes closed (EC). A total of six pertinent ROIs were chosen: anterior cingulate cortex (ACC), posterior cingulate cortex (PCC), left and right inferior parietal lobule (LIPL and RIPL, respectively), and left and right dorsolateral prefrontal (LDLPF and RDLPF, respectively). Although activity was originally recorded at the level of the scalp (with a total of 61 electrodes), the software LORETA was employed in order to solve the inverse problem transformation, thus, extracting the data necessary for the analysis of the referred ROIs at the level of grey matter voxels inside the cortex. During EO higher delta (maximum at 3 Hz) and beta activity (from 23 to 28 Hz) in frontal areas was observed, whereas during EC a prominence of alpha activity in the occipital cortex (maximum at 10 Hz) was present. The iCOH analysis of EO and EC revealed that the posterior cingulate cortex was a major sender of mainly alpha oscillations to all other ROIs. Interestingly, during eyes open, this function was turned off, and the anterior cingulate activated the dorsolateral pre-frontal cortices (mainly theta-alpha oscillations, Figs. 5.11 and 5.12).

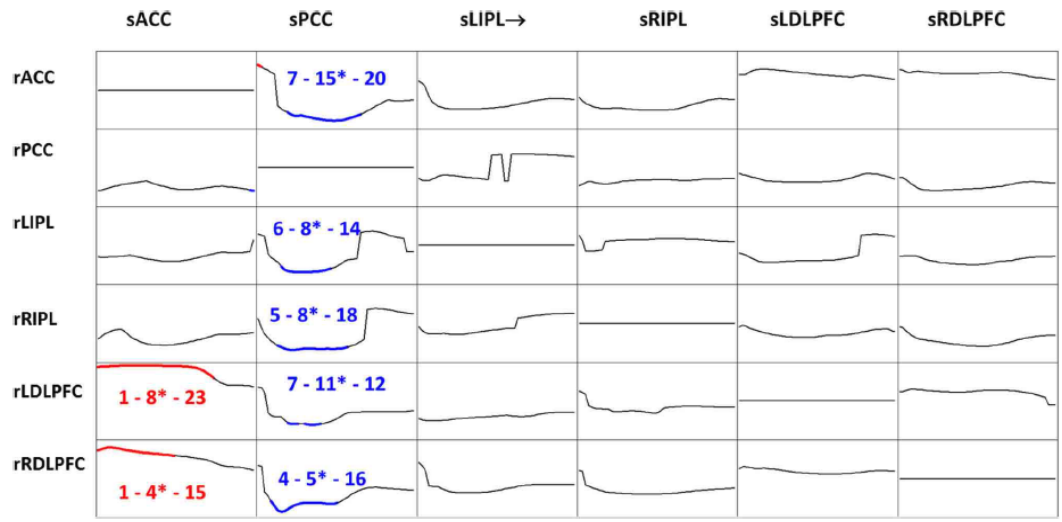


***Fig. 5.11.*** *Application of iCOH to the conditions EO (eyes open) and EC (eyes closed). Differences in iCOH between the conditions are illustrated. Columns are senders (indicated by the prefix s before the name of the ROI) and rows receivers (indicated by prefix r). Frequency axis from 1 to 30 Hz. Color lines indicate the frequency range of statistical significance (P<0.05), with red for EO>EC and blue for EO<EC in effective connectivity. Numerical values indicate the lower and upper limits for the frequency ranges of significance, with the middle number corresponding to the frequency where maximal statistical significance is achieved. Abbreviations in the figure: ACC: anterior cingulate cortex. PCC: posterior cingulate cortex. LIPL and RIPL: left and right inferior parietal lobule LDLPFC and RDLPFC: left and right dorsolateral prefrontal cortex.* Reproduced from (Pascual-Marqui et al., 2014a).

This paper (Pascual-Marqui et al., 2014a) also contains a section with simulated EEG data to compare iCOH and PDC. It demonstrated that iCOH was superior than PDC. In this regard, two particularly prominent features were described: first, although both methods are theoretically normalized to be in the range 0 to 1, the PDC shows in practice a much smaller range of values than iCOH. As it was indicated previously, this problem stems from the way the PDC is normalized. Its values decrease when the number of channels increase, but not so for the iCOH, explicitly defined to overcome this obstacle. Secondly, the PDC exhibits in general a more precise frequency response, especially for the localization of local optima (useful to pinpoint the

frequency values where transmittance of a causal link reaches either local maxima or minima).

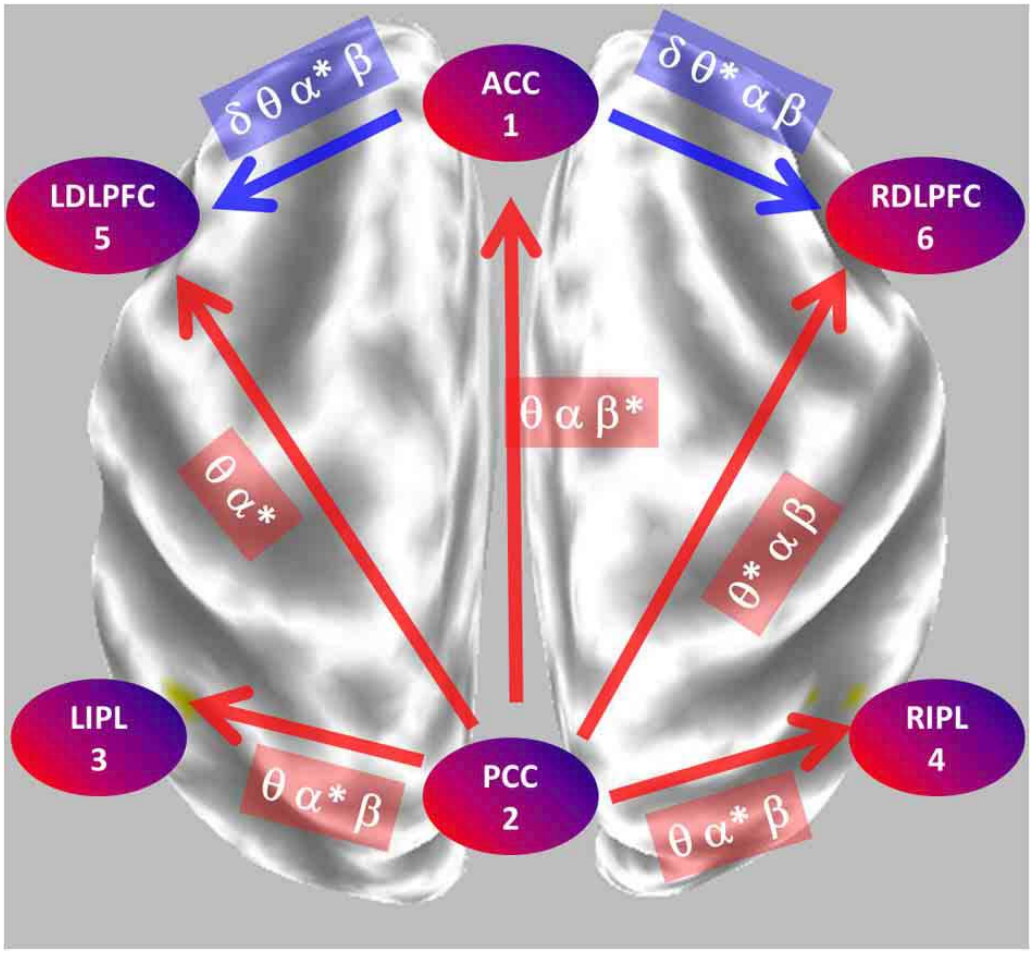


***Fig. 5.12.*** *iCOH analyses illustrating connectivity during EO and EC. Blue arrows correspond to the EO condition, red arrows to the EC condition. Greek letters linking ROIs refer to statistical significance of transmission at those frequency bands, where the asterisk indicates the existence of a global maximum for that particular frequency band. Abbreviations in the figure: LDLPFC and RDLPFC: left and right dorsolateral prefrontal cortex. LIPL and RIPL: left and right inferior parietal lobule. ACC: anterior cingulate cortex. PCC: posterior cingulate cortex.* Reproduced from (Pascual-Marqui et al., 2014a).

Finally, Table 5.2 aims to serve as a summary, describing the main effective causality techniques previously discussed in this review. For each technique, the most remarkable pros (advantages, easy practical implementation, efficiency, etc.) and cons (caveats, limitations, flaws, etc.) are listed. On the other hand, Table 5.3 refers to applications of these techniques in the sleep field and mentions some papers where more information can be found. In addition, Table 5.3 contains a brief summary of the main results obtained.

| **Technique** | **Pros** | **Cons** |
|---|---|---|
| Granger causality (GC) | -Linear (based on MVAR model), mathematically and conceptually simple and elegant. In practice, performs very similar compared to other non-linear methods for directionality detection, which are however considerably more complex and mathematically more difficult than GC.<br>-Easy computational implementation; existence of many GC packages freely available online (widely used in econometrics since decades).<br>-Requires less data (e.g., number of epochs) than non-linear methods for proper detection of directional causal interactions. | -Requires previous estimation of order *p* in the MVAR model to produce optimal results.<br>-Analyses are pairwise, and therefore GC neglects the possible existence of third party "confounding variables" directing the observed couplings; therefore, even a highly significant GC index does not guarantee only by itself "true" causality (i.e., direct influence).<br>-Analyses in frequency domain are also mathematically possible, but in practice result very demanding and more difficult to carry, especially if the number of channels is large (DTF much more suitable in this regard, which automatically allows the inclusion of the whole network). |

| Technique | Pros | Cons |
|---|---|---|
| Information theory (IT) | -High order correlations, possibly undetected by other classical techniques (e.g., coherence), can be better captured. Thus, if one deals with EEG data that may include non-linear segments, MI is a good option to include, performing better than classical linear methods at detecting "hidden" non-linear correlations (and of course, is also able to capture linear correlations).<br>-Estimations of directionality are possible with minor mathematical modifications (so called "directed information"), despite of originally not being a method specifically designed for causality; capability of detection and discarding of confounding variables masking authentic causality, when applied to multiple channels (through "partial mutual information"). Hence, MI is considered in the literature a very versatile method (cross-mutual information, partial mutual information, auto- | -Due to its non-linear character, requires much more data (e.g., epochs) than linear methods in order to be mathematically reliable.<br>-In the classical approach, requires the construction of a histogram for the distribution of probabilities; thus, the number of bins to be chosen there exerts an impact in the final output value that can moderately bias the results either positively or negatively.<br>-Although a generalization for arbitrary number of channels (networks analyses) is theoretically perfectly applicable, in practice, the computational load increases very rapidly with the number of channels, thus rendering it only useful for pairwise connectivity (perhaps, up to 3 or a few more channels with a clever implementation). |

| | | |
|---|---|---|
| | mutual information, directed information, etc.).<br>-Can also be applied to a single channel ("auto-mutual information"), yielding a function that generalizes and improves autocorrelation into its non-linear version, which can be used to estimate degree of predictability in the signal (i.e., "chaotic" behavior) for each channel. However, this type of application is not related to connectivity. | |
| DTF/PDC/iCOH | -Allows the connectivity analysis of a whole network without the limitations of pairwise analyses (such as with Granger causality). Thus, computation needs to be run only once, whereas pairwise dependent method increase quadratically with the number of channels. Many authors claim DTF vastly outperforms GC.<br>-Other various similar variants such as iCoh ("isolated effective coherence"), dDTF ("direct Directed Transfer Function") and | -Sometimes, DTF can commit type I error (i.e., detect correlations that are only there via the existence of indirect pathways).<br>-Temporal analyses not included, works only applied and restricted to the frequency domain. Besides, DTF frequency analyses may be computationally demanding.<br>-Values of effective connectivity with DTF/PDC between a pair of channels can be artificially reduced by raising the total number of channels recorded, and vice versa (yet, |

| | | |
|---|---|---|
| | PDC ("Partial Directed coherence") have been proposed to overcome its original limitations. -Mathematically is based on a MVAR approach, which has been proven to outperform the bivariate AR models. In other words, predictive accuracy in the linear model is improved simply because it considers the presence of all available channels (not just a specific selected pair). | always bounded to the 0 to 1 interval); this flaw is not present when introducing the iCOH. |

***Table 5.2.*** *Pros and cons of effective connectivity methods.*

| **Technique** | **Applications in the sleep field** | **Article** | **Main results** |
|---|---|---|---|
| Granger causality (GC) | -Directionality maps during stage 2 of NREM.<br>-Possibility of quantification of degree of asymmetry in the cortical interactions. | "Inference of Granger causal time-dependent influences in noisy multivariate time series". (Sommerlade et al., 2012)<br>"Granger causality analysis during of steady-state electroencephalographic signals during propofol-induced anesthesia" (Barrett et al., 2012) | -Bidirectional causal flow between ACC and PCC with LOC, although greater in the PCC to ACC directionality. Although statistical differences were present in all bands, they are more prominent in the high frequency range. |
| Mutual information (MI) | -Effective connectivity with the aid of transfer entropy.<br>-Auto-mutual information of single channels in different sleep phases, comparing young versus elderly women.<br>-Cross-mutual information to rigorously assessing functional connectivity (including non-linear correlations). | "Mutual information analysis of EEG signals indicates age-related changes in cortical interdependence during sleep in middle-aged versus elderly women". (Ramanand et al., 2010a) | -Significantly lower values of MI in the elderly group for both NREM and REM sleep in the low frequency range, especially for central areas. MI values rose gradually with increasing sleep depth, reflecting greater cortical interdependence. |

| | | | |
|---|---|---|---|
| DTF/PDC/ iCOH | -Effective connectivity maps and its dynamical evolution during sleep stages (all channels being simultaneously analyzed).<br>-Characterization of cortical networks principally involved in the sleep onset transition.<br>Interactions cortex-hippocampus. | "Inference of Granger causal time-dependent influences in noisy multivariate time series". (Sommerlade et al., 2012)"Topographic analysis of coherence and propagation of EEG activity during sleep and wakefulness". (Kaminski et al., 1997)<br>"Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance". (Kaminski et al., 2001)<br>"Changes in fronto-posterior functional coupling at sleep onset in humans". (De Gennaro et al., 2004) | -During slow wave sleep, PDC shows a reversal in the direction of causal flow between the cortex and the hippocampus (dentate gyrus). The causal reversal takes place during spindle episodes (an average of 2 s), from hippocampus to cortex.<br>-DTF showed a prevalence of occipital to frontal transmission (particularly in the low frequency range) preceding SO, whereas after the transition, the directionality pattern was inverted (frontal to parieto-occipital, significant for all frequency bands). |

***Table 5.3.*** *Applications of directed connectivity estimations in sleep recordings.*

In this review, we have seen the main distinctions between functional and effective connectivity (often misinterpreted in the literature) and have concisely explained the inner workings of a list of prominent methods allowing the successful analysis of brain connectivity by using EEG/MEG. It includes, among others, a thorough branch of methods stemming from information theory (cross-mutual information, transfer entropy, etc.) and multivariate autoregressive modeling based methods (DTF, PDC, etc.). The techniques described in the review can be used either to quantify the strength of coupling of synchronous activity (i.e.., functional connectivity) or the causal flow affecting transmission of information between different brain regions (effective connectivity), moreover, offering the spectral components in which the influence is particularly high (or low). Main caveats of implementation and advantages (at the theoretical or practical level) were also discussed. A special emphasis was given to the applications in the sleep field, which were discussed whenever possible using each of the reviewed techniques as an example where its use can be seen in action. Yet, the methods are totally general and can be extended to almost any other context in the computational neurosciences with fruitful results.

# 6. Discussion

*"Science is much more than a body of knowledge. It is a way of thinking; a way of skeptically interrogating the universe with a fine understanding of human fallibility."* Carl Sagan[6.1]

## 6.1. Limitations: practical and theoretical

We also need to describe several limitations of our studies, i.e., both conceptual and empirical factors that are very likely to have exerted an influence on our results, and that if improved for future studies, could make our main claims more robust.

### 6.1.1. Number of EEG recording channels and relationship with LORETA

Firstly, we need to comment on the issue of the number of electrodes, which is the starting point for the whole analysis. We used 27 EEG channels, probably sufficiently large; yet, this selection is not optimal, as a higher number of electrodes would have yielded better results. However, the relationship between the number of channels and the final precision of the results in the problem of source localization is not as straight as it might appear (Pascual-Marqui, 2002). In LORETA, the spatial resolution of the inverse problem (not referring to the number of voxels employed by LORETA, but to the variance in the localization of generators, with a range of error of approximate spherical symmetry) does indeed initially get better with increasing number of channels, until it saturates in an asymptotic like fashion (Pascual-Marqui et al., 1994, Pascual-Marqui et al., 2002). This saturation happens when reaching about 64 channels (therefore, high density EEG with 128 channels or more will not produce a resolution twice as good as 64 channels, but approximately the same) (Pascual-Marqui et al., 1994). Nevertheless, at the same time, the effect of noise in the recording channels builds up linearly with its number (that is, the noise function is additive) (Lantz et al., 1997). The undesirable effect of noise cannot be simply circumvented, as LORETA maintains the

GIGO principle (i.e., “garbage in” as input produces “garbage out” as output), although LORETA also offers the possibility of regularization for the pseudoinverse matrix, that leverages to some degree this problem (Pascual-Marqui et al., 2002). In addition, smoothing filters applied to the EEG data would increase the signal-to-noise ratio (SNR), but at the expense of progressively reducing valuable information as well when introduced to LORETA (Seeck et al., 1996, Esslen et al., 2008b). On the other extreme, a number of EEG channels lower than 19 would produce negligible noise effects, but the spatial resolution would be so poor that we could not conclude any reliable results without skepticism (Pascual-Marqui, 2002). Putting all these variables together, we get two trends: better resolution with increasing number of channels, but also worse quality (at a given moment, rendering useless the data, given the low SNR) (Pascual-Marqui et al., 2002). Therefore, we can conclude that more than 27 electrodes would have yielded better results for the inverse solution, however, we cannot directly recommend high-density EEG for future studies due to the previous considerations, in which a compromise needs to be found. Finally, it should be considered that with a high number of electrodes, the caution for getting its corresponding positions as accurate as possible must be raised, as the LORETA solution will become increasingly sensitive to the available EEG information (in contrast, for a low number of electrodes, the precision for the spatial coordinates is not as important since, anyway, the LORETA solution will exhibit poor spatial resolution) (Pascual-Marqui et al., 1994).

### 6.1.2. Number of participants and statistics

Secondly, we must comment on the number of participants in our recordings, another important variable that affects the results. We only used eight young healthy subjects, which is a rather low number, and definitely should be increased in further studies. On the contrary than with the first point, here we can be sure that a larger amount of this variable will always improve the quality of the statistics and render more reliable results. However, we applied the method of statistical non-parametric mapping (SnPM) (Nichols and Holmes, 2002). SnPM belongs to the family of “bootstrapping” methods, which are generally

characterized by being able to flexibly adapt even when the number of samples is rather low (Connemann et al., 2001). Indeed, bootstrapping methods were invented for tackling similar issues of insufficient sample size appearing in different sciences; e.g., when obtaining more data is not easily possible (as by costs or time), bootstrapping methods perform a "resampling" of the original data, thus, increasing statistical accuracy (Nichols and Holmes, 2002). In our case, we introduced a very large number of surrogates (5,000), to guarantee maximal precision for the final statistical maps based on *t*-values (Pascual-Marqui et al., 2002). Additionally, bootstrapping methods are also known to be asymptotically exact and convergent, therefore, when increasing the number of subjects, the use of SnPM method can be maintained as well in further studies, even if we had access to a very large sample of subjects (rendering excellent confidence intervals) (Pascual-Marqui et al., 2011). Another advantage is that SnPM does not require to assume any specific theoretical form for the underlying distribution of the studied variable (in our case, *t*-values representing statistical distance between conditions, e.g., baseline and recovery sleep), but is able to infer the true statistics (being asymptotically exact) for any given empirical distribution (Pascual-Marqui et al., 2014b). Hence, SnPM adapts itself to each problem by being distribution independent, rather than applying a "one-size-fits-all" model chosen by the researcher for all necessary statistics, which is prone to fail (e.g., if its assumed a given distribution for a variable that is indeed distributed in another fashion, therefore, rendering distorted results because of false assumptions; thus, the safest policy is "no assumptions at all") (Pascual-Marqui et al., 2014a). Thus, we recommend increasing the number of subjects, but retaining the same methodology that we followed for the subsequent statistical analysis.

### 6.1.3. Use of Magnetoencephalography (MEG) and comparison with EEG

Thirdly, it would be interesting repeating this study, instead of by EEG measurements, using MEG. There are three main reasons to follow this recommendation. Besides verification with another brain recording technique, the MEG gathers complementary information to the EEG (Houck et al., 2017). As we know, the EEG is sensitive to cortical gyri, whereas sulci are rather neglected, given the cancellation of opposite currents by geometrical considerations (Lantz et al., 1997, Srinivasan, 2006). On the other hand, being magnetic fields orthogonal to electric fields, the opposite is true for the MEG, that is, the MEG picks information that the EEG does ignore, and vice versa (Ewald et al., 2013). Hence, source localization with MEG would not only serve to check our main conclusions, it could also even extend them (adding new information). Second, the MEG is reference absolute, and thus, it is not necessary to bother with the right election of a reference channel, which can introduce artifacts, thus, rendering the MEG as "cleaner", although in the case of LORETA, this issue is only of moderate importance given the fact that source localization starts by re-referencing all data to the average reference (Franzen et al., 2013, Houck et al., 2017). Furthermore, the MEG typically presents very high SNR, so data quality can easily surpass EEG measurements, dealing with much less noise fluctuations (Schoffelen and Gross, 2009). In addition, as the technology of SQUIDs (sensor channels of the MEG, dependent on superconductivity) improves with time (e.g., with lower temperatures and superconductive materials, further improving SNR), so will do the accuracy of the measurements (a fact that cannot be said about EEG, based on a different physical principle, potential differences, and where thermal noise of resistors is much harder to reduce compared to MEG) (Schoffelen and Gross, 2009). As the third main reason, the magnetic susceptibility of the brain is close to zero (so volume conduction is of much lower importance than with EEG), and hence, source localization is expected to yield very good results, as electromagnetic distortions (e.g., as introduced by the scalp) would be minimal (Srinivasan, 2006, Sharon et al., 2007). The only caveat is that the MEG, although it would produce more accurate

source reconstruction maps than the EEG for cortical sources, is also slightly worse than the EEG for deep sources (as typical brain magnetic fields are extremely weak and, furthermore, rapidly decay with distance, faster than electric fields), yet, in any case LORETA is less reliable under these conditions (Schoffelen and Gross, 2009). Of course, if possible, we also recommend the combination of EEG and MEG, which would sensibly improve the accuracy of source localization thanks to their complementary nature, although in this case correction of artifacts also becomes more difficult due to electromagnetic coupling of fields (Sharon et al., 2007). We also have to take into account that with current available technology, MEG recordings are more difficult to carry out at the experimental level than EEG recordings when subjects are put to sleep.

### 6.1.4. iCOH: criticism and effects when combined with LORETA

The final consideration is of more theoretical nature. Even when LORETA is the best available source localization method (as compared to other methods, such as weighted norm minimization or dipoles, which fail even for moderately simple source distributions) (Wan et al., 2008, Jatoi et al., 2014), given that is has true zero localization error, we cannot neglect the blurring of its solution, that increases as sources deepen in subcortical regions of the brain (Grech et al., 2008, Pascual-Marqui et al., 2014a). It must be pointed that the blurring introduced by the LORETA method not only affects power spectral density maps, but also connectivity results (Jatoi et al., 2014). This unfortunate effect is a consequence of the fact that, even if discounting electrical volume conduction (a phenomenon that any moderately good source localization method always tends to mitigate), LORETA produces some mixing in the time-series solution for a given voxel, which is “contaminated” by nearby voxels (especially for deep sources) (Pascual-Marqui et al., 2002). This blurring comes from the very mathematical formulation of LORETA, which minimizes the Laplacian of source currents, and somehow, constitutes the price to pay for its zero localization error (Del Felice et al., 2014). Therefore, if signals at the source level become intertwined or mixed, connectivity

results are more prone to be dubious, and additionally, the issue of causality and directionality becomes even harder to disentangle (Srinivasan, 2006, Colclough et al., 2015). In order to circumvent this problem, we used a lagged effective connectivity measure, the lagged iCOH (other lagged effective connectivity techniques could also be employed, that is, if using a linear approach, by pruning the zero-order covariance matrix of the multivariate autoregressive fitting) (Pascual-Marqui et al., 2014a, Bastos and Schoffelen, 2015). The particular selection of iCOH is due to its vast theoretical superiority, as was exposed in chapter 2. In contrast, the instantaneous iCOH, without further corrections, would have introduced some mixing due to both the nature of the inverse solution provided by LORETA and electrical volume conduction effects, so we omitted it and turned directly to the lagged version (Pascual-Marqui et al., 2014a). Nonetheless, counteracting these caveats, we must also state that we did not use a high number of channels (so volume conduction is present, but not as patent as with high-density EEG) and that our defined regions of interest were always several centimeters apart (so both previous effects were negligible, but in other experiments could easily show up, where a lagged connectivity version would better prove its superiority) (Kaminski and Blinowska, 2014).

However, even if the lagged iCOH is a better option than the instantaneous iCOH, the former election is still not optimal. Recently, an improvement of the iCOH has been introduced, that allows tackling the issue of potential signal mixture without “severing” valuable information from the network contained in the zero-lagged covariance matrix of the system (as the lagged iCOH does), called innovations orthogonalization (Pascual-Marqui et al., 2017). A brief summary of the workings of this method is next exposed. To begin, this approach is more realistic and rigorous in the way it treats the zero-order correlation matrix, finding a compromise between a total neglect of its components (as with lagged iCOH) and inclusion, though with no correction at all (as with instantaneous iCOH), thus, finding an equilibrium that only retains the best of both approaches. The rational for introducing this new method is based on the fact that, in general, it is an error to assume that zero-order cross correlation elements are only

intrinsically related to volume conduction effects (and in the case of LORETA, blurring of the inverse solution) (Colclough et al., 2015). Thus, even if the signals are "unmixed" (that is, not mutually contaminated), zero-order correlation elements might still have statistically significant values greater than zero (in absolute value), that is, instantaneous effects might be present even without assuming blurring or electrical volume conduction. A simple case to illustrate this point is to imagine a source region that propagates information towards two brain areas (called *X* and *Y*), taking the same time to do it (e.g., both are located at the same distance from the source, being fixed the velocity of transmission of information through white matter bundles). We would have that in such a case, *X* and *Y* would have significant zero-order correlation matrix that is entirely independent from assumptions of electrical properties of the brain or source localization pitfalls, as it is simply based on the geometry of the network. Consequently, this example motivates the search for a method that, while correcting for false zero-order correlation elements (as due to the systematic error sources previously explained), still retains the "true" elements of the matrix, that reflect the intrinsic characteristics of the network (of course, it is not necessary that these "true" elements have always to exist, but experimentally, this is usually the case) (Pascual-Marqui et al., 2017). The gist idea behind this method is that, even for totally unmixed signals under idealized conditions, such non-zero elements will, in general, continue to exist, so innovations orthogonalization applied to iCOH aims to "separate the chaff from the wheat" once and for all.

In contrast, applying direct "zeroing" of the cross-correlation matrix to non-diagonal elements (as occurs with the lagged iCOH), called "leakage correction", is not a proper solution, as spurious connectivity would imply significant non-zero values, but the opposite claim is not true, hence, constituting a biased or flawed sort of correction (Colclough et al., 2015). Furthermore, leakage correction methods could even worsen the problem they intend to solve, as in the hypothetical example of totally unmixed signals, these methods would erase true information from the system (all instantaneous connections)

of true biological origin, when it should leave the system invariant (as, by design, no artificial instantaneous connections were introduced) (Pascual-Marqui et al., 2017). Hence, applying corrections just for the sake of it, without understanding the architecture of the system, it can introduce misleading unintended consequences, because from the perspective of leakage correction methods, we are left blind to the fact that non-zero elements are distinctive, as they can have either a biological or artificial nature. Thus, paradoxically, in the previous example, applying no corrections at all would have actually rendered the real connectivity map of the network, a criticism that can always be directed at the lagged iCOH (that is, destroying actual instantaneous connections not spuriously generated) instead of the innovations orthogonalization correction to iCOH (especially for complex networks and a large number of recording channels). Given that both instantaneous and lagged iCOH techniques are, in a rigorous sense, flawed and incomplete methods, innovations orthogonalization aims to correct all deficiencies, as it identifies biological signals appearing in the zero-lag cross-correlation matrix and is able to "unmix" even strongly mixed signals (Pascual-Marqui et al., 2017). Thus, as compared to the instantaneous iCOH, mixed signals are unmixed, and as compared to the lagged iCOH, it performs a correction that is strictly specific to mixed signals, while respecting zero-lag elements of true biological origin (whereas the lagged iCOH would be like a doctor that makes amputation also to healthy limbs). We conclude this section by mentioning the introduction of a recent mathematical method, called structured sparsity, which by using a hierarchical Bayesian formulation with certain algebraic constraints for the current norms is able to improve the spatial resolution of deep sources (Krishnaswamy et al., 2017, Paz-Linares et al., 2017). Hence, by dealing better with the mixing problem previously described, we propose that future studies should combine LORETA, as corrected with the structured sparsity constraint, and the iCOH or another similar effective connectivity method, as corrected by the innovations orthogonalization.

## 6.2. Future directions to explore

*"In every discipline, progress comes from people who make hypotheses, most of which turn out to be wrong, but all of which ultimately point to the right answer."* Milton Friedman[6.2]

The topic of sleep is very rich and crucial for a thorough understanding of the brain, hence, it is tightly coupled with other fascinating areas of research in neuroscience: memory, dreaming, synaptic dynamics, etc. (Born et al., 2006, Cirelli, 2013, Aguirre, 2016). Consequently, by attempting to link sleep research with these other topics, due to shared "mutual information", investigation is forwarded in all related fields. Here we want to make connections with some not so distant fields, which could benefit symbiotically from sleep research: creativity and the neuronal correlates consciousness (NCC). We will also discuss how the SO process that we investigated in a standard case could also be tweaked as a function of, e.g., personality traits normally distributed on the population, and the rationale for this assumption. We want to end this introductory section by recommending the research of sleep onset not only for the forebrain and midbrain, but also for the cerebellum (where research is much more scant), as the cerebellum shares extensive reciprocal connections with many brain areas, in order to get a complete picture (e.g., new evidence connects sleep disorders, such as REM sleep behavior disorder and chronic insomnia, with cerebellar malfunctioning, and vice versa) (DelRosso and Hoque, 2014, Canto et al., 2017).

### 6.2.1. Sleep and creativity

*"If a solution fails to appear, and yet we feel success is just around the corner, try resting for a while. Like the early morning frost, this intellectual refreshment withers the parasitic and nasty vegetation that smothers the good seed. Bursting forth at last is the flower of truth."* Santiago Ramón y Cajal[6.3]

Regarding the association between creativity and sleep, and more particularly, the sleep onset transition, we must bring the topic of hypnagogic hallucinations, and the fact that many artists and scientists throughout history have used this technique to boost their creativity (e.g., Thomas Edison, Nikola Tesla, Salvador Dalí, etc.) (Jimenovaldes and Jimenobulnes, 1993). A large survey ($N$=13,057) found that the prevalence of hypnagogic hallucinations in the general population is 24.8% (in contrast, hypnopompic hallucinations, occurring when awakening, are more rare, with 6.6% of population) and that they are progressively less frequent with aging (Ohayon, 2000). A fMRI study indicated that brain areas showing increased blood flow during hypnagogic hallucinations include the ventromedial prefrontal cortex, frontal eye fields and hippocampus (hence, autobiographical memory retrieval could be used as raw material to form these hallucinations), while the thalamus and posterior cingulate cortex exhibited a decrease (consistent with external disengagement and reduction of consciousness in such drowsy state) (Ong et al., 2015). When undergoing such a hallucinatory state, cases of transient amnesia (subjects must be awaken by the researcher for noticing it) have been frequently reported, particularly concerning events prior to sleep onset, which might be due to a hippocampal to cortical functional disconnection, thus, producing an impairment in short-term memory (indeed, the hippocampus shows spindle activity preceding cortical regions, hence, manifesting earlier signs of sleep) (Sarasso et al., 2014).

It has been reported that people scoring highest on creativity measures experience more frequent and vivid mental imagery in the form of hypnagogic hallucinations during the sleep onset transition compared

to the general population, so it is likely that they are among those who could benefit the most (Watanabe, 1996). This phenomenon can be attributed, at least partially, to the emergence of alpha activity during the SO transition (as shown in chapter 3 regarding the results of this band, whose activity significantly increased in prefrontal regions). In this way, there is robust evidence that EEG alpha activity is sensitive to creative ideation, and that there is a positive correlation between individual creativity, originality of ideas and alpha power (Fink and Benedek, 2014). A neurophysiological explanation for this relationship is that alpha activity prevents bottom-up processing (as it is reasonable to expect during sleep onset), so in the absence of bottom-up stimulation, top-down activity can be more effective, e.g., by recombination of unrelated semantic information and other memory processes (autobiographical, etc.) (Benedek et al., 2014). Additionally, it has also been experimentally proven through 10 Hz transcranial-alternating current stimulation applied to the prefrontal cortex that creativity levels can be significantly boosted in a laboratory session (Lustenberger et al., 2015). The authors argue that inhibitory top-down control benefits creativity, which is connected to the external disengagement (that is, at the neuronal level, a form of inhibition, as the environment becomes a “task-irrelevant process”, so to speak) happening during sleep onset, while there remains still certain conscious awareness (Esser et al., 2009). Furthermore, during the SO transition, norepinephrine and cortisol levels are known to undergo an overall reduction, and it has been experimentally shown that lower concentration of these hormones benefit creativity, possibly due to a modulating effect on cognitive control and cortical arousal in prefrontal regions (Kuhn et al., 2014).

Nonetheless, the question of the neuronal basis of creativity is still only partially understood, e.g., it is known to be connected to the fronto-striatal dopaminergic system (striatal dopamine relates to flexibility in searching new patterns, while the prefrontal cortex is important for perseverance and evaluation) and that there is an inverted U-shaped relationship between dopamine levels and creative performance (Boot et al., 2017). Besides purely dopaminergic regions (such as the ventral

tegmental area), creativity (as measured by a divergent thinking task) has also been related to gray matter volume differences in the dorsolateral prefrontal cortex (the most important hub of the task-positive network), the precuneus and adjacent posterior cingulate cortex (key areas of the DMN), as assessed through voxel-based morphometry (Takeuchi et al., 2010). For instance, gray matter volume in the precuneus is correlated to visual and mental imagery, linking hypnagogic hallucinations with creativity (Dang-Vu et al., 2008, Franzen et al., 2013). In addition, gray matter volume in the precuneus relates to proneness towards inwardly directed attention (Franzen et al., 2013). The previously mentioned areas were also defined as regions of interest in our own study, and their relevance was proved to be prominent (in particular, for the posterior cingulate cortex/precuneus, guiding the SO transition), as shown in chapter 4. Another study on creativity related originality (quality of ideas, required to be both novel and useful) and fluency (the variety and number of ideas produced on a divergent thinking task) to gray matter in the precuneus, further highlighting the importance of this region for creative ideation (although for fluency there were additional neurobiological differences dependent on IQ) (Jauk et al., 2015). Finally, the amount of gray matter throughout the whole DMN positively predicted cognitive flexibility (ability to overcome rigid patterns of thought and make novel associations between concepts), which is a necessary element for creativity (Kuhn et al., 2014).

Hence, the state of “dormiveglia” (Italian word for half-sleep and half-awake) occurring during SO (since, as shown in chapter 3, it is a local cortical process), could serve an ideal state of mind-wandering (thus, less control and more flexibility) for the DMN to break rigid schemas and enhance associations, although it is also expected to show inter-individual variability according the gray matter volume on this network (Jimenovaldes and Jimenobulnes, 1993, Jauk et al., 2015). However, as the brain is plastic, it is possible that through divergent thinking training, inter-individual differences could be reduced by increased gray matter volume on the DMN achieved after regular practice, hence, “unlocking” the benefits of the SO transition for creative production on a wider population. Indeed, creativity has been said by economists and

entrepreneurs (from a global survey of 1,500 CEOs by IBM) to be “the number one skill for the 21st century” in order to remain competent and prosperous, and perhaps (although this remark is debatable), it constitutes the only human trait that cannot be replaced and automatized by progress in robotics anytime soon (Bottani, 2010). In summary, for all these reasons, we speculate about an interaction between creativity and possible inter-individual differences in the specific unfolding of the SO transition and how there exists a common ground or intersection between these two fields (i.e., sleep and creativity) (Marguilho et al., 2015, van Heugten-van der Kloet et al., 2015, Llewellyn, 2016, Ram-Vlasov et al., 2016) . As a final example, low cortical arousal associated with NREM sleep could enhance to ability to access remote associations (a key component of creativity), with a study showing enhanced fluency and flexibility appearing during stage 1 of NREM sleep and boosted originality linked to stage 4 of NREM (Drago et al., 2011). In this discussion, we intentionally left the topic of dreams and creativity (as famous benzene structure example by Kekule) (Strunz, 1989) as it is outside the SO transition, which on itself, makes the case of a relationship between creativity and sleep for future studies even stronger (Hobson and Friston, 2012, Tsoukalas, 2012).

### 6.2.2. Sleep and consciousness

*"All men whilst they are awake are in one common world: but each of them, when he is asleep, is in a world of his own."* Plutarch[6.4]

We also want to stress the connection between the study of sleep and consciousness (Hobson and Friston, 2012, Siclari et al., 2017). A common way to find out how a highly complex system operates is by analyzing cases when it fails. This method can be found across many sciences and engineering (e.g., knock-out animals in genetics), but takes more prominence in the study of the brain (e.g., by studying brain lesions, it could be discovered the anatomical position for Broca's area and its purpose) (Lorch, 2011, Koutroumpa et al., 2016). In our opinion, the case of sleep can be potentially used also as "a royal road to the conscious" (paraphrasing Freud) (Hegener, 2005), because, if during SO, consciousness gradually fades away, studying this process in more detail could be an extremely powerful method to unravel the specifics underpinnings of consciousness (e.g., the important role of the thalamo-cortical loop). One way to tackle the question of NCC that complements functional and effective connectivity based approaches is by using graph theory (Bullmore and Sporns, 2009, Sporns and Betzel, 2016). Graph theory offers great potential to the neurosciences; if connectivity methods in neuroscience measure the strength and directionality of the edges of a network, graph theory goes to a deeper, abstract level, by analyzing its underlying topological architecture: modularity, small-world properties, centrality, etc., and it already benefits from centuries of applicable theorems and interesting coefficients established by mathematicians (He and Evans, 2010, Huang et al., 2016). Additionally, graph theory has been successfully applied to other major problems in neuroscience, e.g., to predict differences in fluid intelligence based on small-world properties (characterized by short path-length and high clustering; the more small-world features on fronto-parietal networks, the higher IQ) and to detect early signs of neurodegenerative diseases (Alzheimer, Parkinson, etc.) (Langer et al., 2012, Dipasquale and Cercignani, 2016). A recent study performed by Siclari et al. sheds new light on the NCC (Siclari et al., 2017). They confirmed that dreams are not limited

to REM sleep, but also NREM sleep as well. When dreaming activity occurred (a purely inward form of consciousness, although interestingly, the same brain areas were recruited for dream fantasies as for daily external activity), there appeared several neuromarkers specific to this state (as compared to no dream experience): high beta activity and reduction of delta/theta activity. Interestingly, an in agreement with our research (see chapter 4), this neuromarker was best located on a posterior "hot zone" matching the precuneus/posterior cingulate cortex, which corroborates our thesis of the crucial role of this subcortical region for consciousness and sleep onset (guiding external disengagement by shutting down cortical areas as the DMN breaks down, being its most important hub) (Vogt and Laureys, 2005, Amico et al., 2014). The posterior cingulate cortex is crucially involved in self-referential processing and, besides sleep onset (where the sense of self fades away) as we proved with iCOH, deactivation of the posterior cingulate has also been reported during meditation (as meditators must suppress self-referential processes) or psilocybin intake (an hallucinogenic reported to lead to ego dissolution, additionally sharing with SO being linked to decreased medial prefrontal to posterior cingulate coupling) (Brewer et al., 2013).

Although research on sleep using graph theory exists (Uehara et al., 2014), indicating reduced nodal efficiency (quantified by decreased path-length between nodes) of most DMN hubs during stages 1 and 2 NREM sleep compared to prior wakefulness (furthermore, corroborating the importance of the DMN for consciousness), little is known concerning SO in the context of graph theory. Hence, more research is needed, e.g.: to know how small-world features gradually evolve with time, how they can be connected to key aspects of consciousness (information integration, external disengagement, hypnagogic hallucinations, etc.) and, finally, how graph theory relates to power spectral density and functional or effective connectivity changes as SO unfolds (as we exposed in this project). A recent study on this topic found that, for the sigma band, frontal networks were less ordered (or more random, reducing path-length and aiding spindle spreading) during SO, while for low frequency bands, networks were

more ordered (thus, allowing fronto-parietal coupling observed along a longitudinal axis, as shown in chapter 4) (Vecchio et al., 2017). Given that whatever shift from small-world will favor either global connectivity over local connectivity or the opposite trend, the tradeoff relationship between slow wave activity (facilitated by network order so neuronal ensemble synchronization can build up and explaining why SO is a local process) and spindle activity (facilitated by network randomness for quick cortico-cortical spreading) (Aeschbach and Borbély, 1993), that becomes more evident during sleep deprivation, can be better understood in light of graph theory). It could also explain the inverted U-shaped pattern we saw in chapter 3 for sigma activity, as when rising of delta activity caused by sleep homeostasis is so high that it dominates over sigma activity, a very ordered state of brain networks (high clustering) will also dominate at the topological level, and consequently, hampering spindle transmission, so there will we be a decline of sigma activity. Reduced small-world topology in delta and theta bands could also relate to the breakdown of effective connectivity we commented in chapter 4 and explain what Massimini et al. experimentally found with TMS (Massimini et al., 2005), as more segregation will hinder any stimuli propagation, and possibly, constituting the reason why the inner sense of conscious identity fades under this network topology (the brain cannot integrate distant areas anymore, it is too much segregated) (Storm et al., 2017).

Another parameter derived from graph theory, modularity (measuring "segregation" and the accuracy for defining clusters), has also been applied to sleep, showing that the sleep deprived brain suffers an overall reduction of modularity (DMN brain regions are "misconnected" for performing a cooperative work, possibly explaining lower cognitive efficiency and mental tiredness), an interesting parameter to track as SO advances (Ben Simon et al., 2017). A brain with less modularity by sleep loss would be less segregated on nodes of its networks, exhibiting more randomness on its architecture and consequently, diverging from a small-world network (which represents the cognitively optimal state). Lower network modularity caused by sleep deprivation also provides an explanation for why the SO transition is more spatially homogenous under this state, as the

network has lost its capability to segregate itself into relatively independent and distinct local modules, and as a result, instead of a modular understanding for the brain networks, a global-centric description takes precedence (De Gennaro et al., 2004, Nir et al., 2011). In general, the brain seems to operate best at a middle point between a lattice type of network (constant number of edges per node repeated in the same pattern) and a random network, so it can benefit from both high clustering and low average path-length (keeping the best properties of both extremes); e.g., neurodegenerative diseases such as mild cognitive impairment are shown to correlate with divergence from small-world features (Dipasquale and Cercignani, 2016). In this way, brain network are capable of being efficient to work in tandem between nearby nodes and specialize in "compartments" (clusters), but without sacrificing long-range flexibility and easiness to share information (short path-length) (He and Evans, 2010, Douw et al., 2014).

We think the previous fact provided by graph theory could explain our finding in Chapter 4 of lower iCOH reductions during the SO transition in the recovery case (after 40 h of prolonged wakefulness): if neuronal networks are already more "disorganized" (as modularity is lower, so cluster arrangements dissolve or blur, being worse defined), it makes sense to need a smaller connectivity reduction for the final and complete breakdown of the DMN. Moreover, it would be reasonable to expect propagation of spindle activity to be hindered in the recovery case due to the lower modularity (as disconnection inhibits transmission), as it was actually revealed through iCOH. To make an analogy, observed differences in SO between conditions would be as observing less time in a race between two objects falling due to gravity, just because in the second case we start from the advantage of an initially lower height (i.e., lower modularity and less small-world features in the recovery case). Thus, SO unfolds quicker in the recovery condition since it requires less reorganization changes in brain topology to occur as compared to baseline (the brain architecture would be preparing itself to be just ready to fall asleep), and consequently, graph theory reveals that the SO transition is "less of a

transition" for recovery sleep (Ben Simon et al., 2017, Vecchio et al., 2017). Another analogy would be a shoelace that is subject to more wear and tear due to a prolonged use, so it is much easier to break when lacing shoes (comparable to the DMN breakdown during SO, since lower modularity produced by sleep deprivation and its excessive entropic build-up renders the network to be loosely defined or "tied").

Supporting this view, when the brain is awake but sleep deprived, resting thalamic connectivity with multiple frontal and temporal regions is significantly reduced, explaining impaired vigilance levels and establishing a "near-to-uncoupled" thalamic state to facilitate the transition (Ben Simon et al., 2017). It is possible that the modularity parameter itself could also be subject to homeostatic regulation (paralleling Process S), with sleep deprivation increasing randomness of brain networks (i.e., lower modularity or higher network entropy than an optimal point) that needs to be reestablished by sleep (particularly, during NREM sleep) (Meisel et al., 2013). This hypothesis could in part explain why the elderly is more prone to neurodegeneration and cognitive decline, as modularity generally decreases with age (He and Evans, 2010) and there is in addition decreased deep sleep (which restores excessive randomness in brain networks) or maybe why chronic sleep deprivation produces hallucinations (schizophrenics show reduced modularity) (Alexander-Bloch, 2014, Vecchio et al., 2017). Furthermore, more randomness in brain networks occurring with sleep deprivation (as revealed by decreased modularity, and consequently, worse segregation) could also explain reduced functional anti-correlation between nodes defining the default mode and central executive networks, a feature found in many neuropsychiatric diseases in the adult brain (the anti-correlation between the two most important networks improves developmentally to optimize brain functioning) (Ong et al., 2015, Dipasquale and Cercignani, 2016). Overall, we think that graph theory and topology are fruitful fields for neuroscientists offering great tools to explore and discover new insights pertaining sleep, consciousness and their intimate relationship, whose potential has not been sufficiently recognized and applied, although it is now beginning to get its

deserved attention (e.g., the Human Brain Project by using algebraic topology) (Reimann et al., 2017).

To conclude, we also propose that since according to one theory consciousness requires information integration (Tononi et al., 2016), it would be useful and interesting to track a parameter for information or complexity content (e.g., Shannon entropy, sample entropy or Lempel-Ziv complexity) (Le Van Quyen et al., 2001, Tosun et al., 2017) as SO develops with time. In this way, in a paper by a group of physicists, application of Boltzmann entropy to different sleep phases revealed that consciousness represents a state of maximum entropy (meaning that the repertoire of possible states the brain can access is as big as possible given its biological constraints, thus, allowing optimal information processing), that NREM sleep is a state of very low entropy and that REM sleep reaches entropy values close to waking (Guevara Erra et al., 2016). Hence, it is possible that the biological evolution of consciousness itself is a manifestation of the second law of thermodynamics. In turn, another study using mice applied the Permutation Lempel-Ziv complexity (PLZC; an improved version, more robust to noise), finding that waking and REM sleep exhibit similarly high PLZC, NREM sleep low PLZC and that with sleep deprivation, NREM sleep (having enhanced slow wave activity) shows even lower PLZC (Tosun et al., 2017). However, these studies only take a stance for sleep phases and lack the temporal evolution approach that we followed in our project. Based on the previous findings, we hypothesize that complexity measures would gradually lower during the SO transition and that the steepness of decay would be larger under sleep deprivation (not only happening quicker, but also descending to lower levels, as slow waves would be more regular during recovery). The fact the recovery sleep shows enhanced slow wave activity can be explained in terms of the brain needing to generate more network order (evidenced by the fact that PLZC shows a decrease in this condition) as a homeostatic compensation for the excess of randomness build up throughout prolonged wakefulness, thus, its function being avoiding hysteresis of brain networks (comparable to a material strongly deformed that later undergoes the opposite deformation to avoid lasting

effects; for the brain, it would avoid cognitive impairments or diseases). We can also argue that, based on a principle similar to Weber-Fechner's law in psychophysics (Johnson et al., 2002), the recovery condition, being already in a highly entropic state, would need stronger or larger stimuli from the external environment for producing a significant change in the network topology (hence, the impaired cognitive performance), thus, sleep is needed to restore the capabilities of the network to better respond to the environment from a lower entropic state (more easily affected by entropic changes due to experience).

Finally, current theories on consciousness (such as proto-consciousness by Hobson, integrated information theory (IIT) by Tononi and Orchestrated-objective reduction (Orch-OR) by Penrose-Hameroff) should be contrasted and refined through sleep research (e.g., Hobson's theory is explicitly built upon his research on dreams and Tononi has contrasted IIT with sleep data) (Hobson, 2009, Hameroff and Penrose, 2014, Tononi et al., 2016). In order to be a valid theory for consciousness, it would be required to provide satisfactory explanations for how every transition shutting it down (sleep, anesthesia, vegetative states, etc.) is even possible and how it may re-emerge (particularly, why it happens during sleep), what are the requirements for a system to be conscious and to describe precisely blurred conscious states (e.g., lucid dreams, hypnagogia, hallucinatory drugs, etc.) (Hobson and Friston, 2012, Tononi et al., 2016). For example, the IIT of consciousness, based on a series of phenomenological axioms from which it tries to deduce the neuronal correlates enabling them (e.g., exclusion of perceptual experience) claims to be such a candidate theory (Cerullo, 2015, Tononi et al., 2016). Furthermore, IIT exhibits the advantage of offering a mathematical formalism allowing to compute the –purported- level of consciousness of a system (a parameter called $\Phi^{MAX}$), proportional to its overall integrated information (Tononi et al., 2016, Kim et al., 2018). While this theory has many supporters, we must also state that IIT has also received criticism on its set of axioms (critically, the axiom of principle of information exclusion, from which its mathematics are built), moreover, it has been argued that integrated

information might be a necessary, but not sufficient condition for producing consciousness, thus, questioning the basis of the theory (there are also philosophical counterarguments related to fading qualia) (Cerullo, 2015).

Hence, we think entropy and information theory measures constitute good candidates to mathematically quantify the degree of consciousness of a system (such as integrated information $\Phi^{MAX}$ in IIT) (Tononi et al., 2016) and how it may vary when transiting to different states (in particular sleep) (Seth et al., 2011). Developing a differential equation for such index (as it varies with time, like in Process S) could help to reveal, among other issues, the "why" question of sleep as we would know the inner dynamics of the system. We hypothesize that the answer is probably related to a process for entropy dissipation derived from information acquisition and processing throughout the day (whose build-up affect and alter the structure of the brain, as revealed by graph theory), given that all previous entropy measures reveal low values during NREM sleep (trying to "reset" the primordial topological properties, decreasing network randomness or entropy that deviates from small-world topology) (Tosun et al., 2017, Vecchio et al., 2017). Thus, a phenomenon not so different from other body organs getting rid of material or physical entropy taken from the environment (e.g., kidneys), only that we would be dealing with information entropy for the brain (moreover, with the aspects considered to be relevant for learning, i.e., for integration into its structure, analogous to free energy absorption when processing food). Nonetheless, being the brain a system for information processing (perception) and modeling (cognition), yet dependent on a physical substrate, network entropy as assessed by information theory and topology would also manifest a material entropic correlate (biochemical waste products produced by information operations), cleaned up during sleep by the glymphatic system for the correspondent material entropy reduction (Mendelsohn and Larrick, 2013). In this line, additional evidence for the brain raising entropy levels through its daily activity is deduced from findings of boosted production of amyloid beta proteins by insufficient sleep (which aims to its removal), consequently, gradually increasing

the risk for neurodegeneration (Ju et al., 2014, Musiek and Holtzman, 2016, Cedernaes et al., 2017, Lucey et al., 2017).

Finally, regarding dreams, they could serve the function of introducing more network entropy in a state that has excessively low entropy (deep sleep), so its simplicity is insufficient to produce consciousness, therefore, requiring internal simulations to prepare the brain to become conscious and alert after awakening. We also hypothesize that small-world properties constitute the "goldilocks zone" for consciousness and healthy brain dynamics and propose that sleep evolved to counteract the tendency of all systems to build up entropy (as all living beings need to do through several ways for survival, avoiding the deterioration of their inner structure, in this case, neuronal networks). This hypothesis would explain why after sleep deprivation and prior to recovery, people experience greater appetite for caloric foods (seeking to import additional external negative entropy to counteract the excess of inner entropy not restored by sleep) (Brondel et al., 2010). Hence, sleep also serves a thermodynamical reason, as a brain able to periodically undergo a state of entropy reduction and subsequent increase would be, over time, much more efficient to increase entropy in the universe compared to a brain that were not able –its own entropy will ultimately deteriorate it until the cessation of its function of "information digestion", so total entropy will stop to increase compared to the other. Additionally, this thermodynamical framework can explain the cooling of brain temperature during deep sleep by noticing that if brain entropy is reduced in this state, in order to satisfy the second law of thermodynamics (free energy decrease or total entropy increase), the system has to decrease its enthalpy (i.e., to release heat, cooling). Our hypothesis can be connected with Friston's theory of the brain as a Bayesian learning system for free energy reduction (where perception and cognition emerge from a thermodynamic variational principle) (Friston, 2010).

### 6.2.3. Sleep and personality research

*"There are two types of people in this world, good and bad. The good sleep better, but the bad seem to enjoy the waking hours much more."*
Woody Allen[6.5]

Sleep can also be connected with personality research simply by understanding personality measures (e.g., as measured by the Five Factor or Big-5 model (Emert et al., 2017), the most preferred by psychologists) as an independent variable affecting a dependent variable (different sleep parameters). There is also a neurobiological common ground between the most important brain areas involved in the SO transition and some personality traits. One example is given by the personality factor in the Big-5 model known as "openness to experience/intellect" (Beaty et al., 2016) (in the remainder, openness for short), a normally distributed personality trait reflecting aesthetic appreciation and a cognitive style tendency to engage in imaginative and abstract processes (openness positively correlates with IQ, being the only personality factor having an association with intelligence, but they are distinguishable) (Beaty et al., 2016, Kaufman et al., 2016). Openness correlates with creativity; therefore, if we talked previously about the relationship between creativity (which can be understood not only as a skill, but as a personality trait), there must also be a relationship between at least aspects of personality related to creativity and sleep (in fact, as we will see, there are very likely much more traits) (Jauk et al., 2015, Kaufman et al., 2016). Applying graph theory, it has been found that openness correlates with topological efficiency (i.e., a short characteristic path-length allowing fast information exchange) of the DMN (explaining 18% of the variance in DMN functioning), with subjects displaying higher openness having shorter average path-length throughout the DMN (probably, being the underpinning for the link between creativity and openness, as an overall shorter path-length makes it easier to draw connections between distant concepts) (Beaty et al., 2016). As we have demonstrated using the iCOH in chapter 4, that the DMN plays a major role in the SO transition. Consequently, we speculate that future sleep

studies could take into account individual differences in openness in relationship to the specific unfolding of the SO transition and perform statistical analyses that would separate effective connectivity dynamics as a function of openness. Given that people scoring high in openness display different topological properties in the DMN, it is reasonable to suppose an influence on the breakdown of the DMN occurring during SO and/or other sleep related features (e.g., given the shorter characteristic path-length of their network, showing enhanced widespread spindle activity and faster cortical diffusion than normal subjects). Finally, as openness relates to imagination and creativity, it is very likely that hypnagogic hallucinations are more frequent among this subpopulation, therefore, they constitute an excellent group of study for improving the neuroscientific understanding of this phenomenon occurring during SO (Jimenovaldes and Jimenobulnes, 1993, Watanabe, 1996).

Another personality trait that could affect sleep is introversion or extraversion, first defined by the psychologist Carl Jung (Jung, 1921), broadly refereeing to preference towards inner or outer forms of stimulation, respectively, and whose preliminary neurobiological correlates were later theorized by Eysenck's model of personality (Kumari et al., 2004, Killgore et al., 2007). There is plenty of evidence that introverts and extraverts have neurobiological differences (they also have genetic differences as this trait has a substantial heritability, around 40-60%), most importantly, a boosted dopaminergic mesolimbic pathway in the case of extraversion (putatively, explaining more propensity towards social rewards) (Smillie et al., 2010, Smillie and Wacker, 2014). Concerning sleep, a key difference between introverts and extraverts relates to tonic arousal of the ascending reticular activating system (ARAS), with introverts (as a population) showing highest cortical and thalamic activity levels (in turn, extraverts have lower resting state BOLD signal in the anterior thalamus, thus, perhaps seeking social external arousal as a compensatory mechanism) (Kumari et al., 2004). The relationship between performance and arousal level is curvilinear (originally proposed by Eysenck on his model based on a previous work by Hebb, but receiving modern experimental confirmation), resembling an

inverted U shape, and when rested, introverts locate slightly to the right of the maximum (with the opposite for extraverts) (Killgore et al., 2007). Moreover, introverts and extraverts also differ in response to sleep deprivation, with extraverts more vulnerable to the effects of sleep deprivation (experimentally shown by exhibiting poorer responses in a battery of psychomotor vigilance tasks or PVT) while introverts were more resilient (in both cases, prolonged wakefulness corresponds to a leftward shifting by fatigue on the curvilinear relationship between arousal and performance, thus, explaining these results under Eysenck's model) (Killgore et al., 2007). On the other hand, it is known that undergoing a reduction in thalamic activity has a detrimental effect over PVT tasks carried under sleep deprivation, whereas thalamic activity is relatively preserved in subjects less affected by sleep deprivation (Ong et al., 2015). Therefore, these results can be interpreted knowing that thalamic deactivation reducing sensory gating begins the process of falling asleep, and given that introverts show higher tonic thalamic activity than extraverts, they will be able to resist for longer periods than extraverts until showing significant decline in PVT performance (and concomitant low thalamic and cortical activity levels). Additionally, extraverts naturally select environments and activities displaying copious sensory stimulation (towards which they are more sensitive at the neurobiological level and are drawn by lower basal arousal), and consequently, they would deplete their thalamic and cortical arousal levels faster than introverts, thus, aiding the SO transition at the end of the day (Johnson et al., 1999, Killgore et al., 2007).

Consequently, we propose further investigating differences in the SO transition between extraverts and introverts pertaining differences in the dynamics of thalamic and hippocampal disconnection preceding cortical signs of sleep. For instance, statistically different time delays between subcortical and cortical sleep onset as shown by first spindle detection, since the strength of subcortical to cortical coupling could possibly vary as a function of the degree of extraversion or introversion (Magnin et al., 2010, Sarasso et al., 2014). Thus, we can hypothesize that introverts would show longer time delays of deactivation for both

subcortical and cortical regions during SO than extraverts due to their higher basal ARAS excitability (however, they would also suffer the downside of more insomnia problems by the same token). We should not forget that SO is a local process, therefore, statistical differences in time delays and latencies between these personality groups would probably reproduce along an anterior-posterior cortical gradient (with frontal areas being the first showing sleep signs, more rapidly causally affected by a more efficient subcortical coupling) (Murphy et al., 2011). Longer sleep latency for introverts could maybe provide higher opportunity for experiencing hypnagogic hallucinations by prolonging sensory cortical sleep onset due to their higher basal cortical arousal (in turn, their frontal lobe has higher tonic blood flow, so it would probably begin resting earlier), thus, temporally stretching this cortically dissociated state (Johnson et al., 1999). Therefore, given what we said earlier in this section, the ideal population for better understanding hypnagogia could be introverts high in openness (not coincidentally, this population will show the most creative characteristics, whose relationship with hypnagogia was previously described).

In general, based on the preceding arguments, we suppose, firstly, that introversion would slow down the SO dynamics described in our project as compared to extraversion (even in a recovery sleep condition after the same duration of prolonged wakefulness, given that introverts show more resiliency, at least if sleep restriction is not chronic) and think that later studies should test this hypothesis. Secondly, we hypothesize that the local aspect of the SO transition would be more salient in introverts (with a clearer anterior-posterior gradient, as their frontal lobe would be subject to greater homeostatic pressure, but not for the sensory cortices), while for extraverts, the SO transition, besides temporally accelerated, would also be more spatially homogenous across the cortex. Hence, we predict a general behavior as if the extraverted brain were slightly sleep deprived compared to the introverted brain during the SO transition, with the latter showing better preserved modularity (a corollary being that for introverts it would be easier to sustain good cognitive performance during late hours). This reasoning at the network level is additionally supported at

the biochemical level by the fact that extraverts would prefer more external physical activities (in contrast, introverts easily overwhelm by overstimulation and are generally more laid-back), hence, their adenosine receptors would have been more stimulated throughout the day, contributing to the proposed result (Smillie et al., 2010, Clark and Landolt, 2017). Thirdly, we also hypothesize that as a consequence of prior brain activity during waking (all these hypothesis hold under conditions of normal daily life), slow wave activity would be larger in the frontal lobe for introverts (at least, some Brodmann areas), while extraverts would have larger slow wave activity in sensorimotor cortices (however, these statistical differences might render, even if significant, only a small effect size). Evidently, the same approach can be applied in other populations in which abnormal thalamic, hippocampal and cortical behavior might be present, thus, possibly modulating to some degree different aspects of the SO transition (e.g., low latent inhibition –reducing thalamic filtering of environmental stimuli due to lower density of dopamine D2 receptors-, synesthesia, sensory processing sensitivity, schizophrenia, dementias, etc.).

There are additional personality factors related to sleep which have also been investigated, such as neuroticism and conscientiousness (Emert et al., 2017). Neuroticism is a personality trait related to propensity to experience negative emotions (in particular, anxiety and rumination) and higher psychophysiological vulnerability to the effects of stress (Ikeda et al., 2015, Nota and Coles, 2018). This population is characterized by insomnia complaints, longer sleep onset latency, less slow wave activity during NREM sleep, more fragmented sleep, less overall sleep duration and lower REM sleep density throughout sleep (Emert et al., 2017). In summary, high neuroticism correlates with all sorts of features reducing sleep quality, which in the long run, would probably establish a bidirectional feedback relationship with this personality trait (Nota and Coles, 2018). On the other hand, conscientiousness, describing responsible, organized, careful, hard-working and task committed individuals, related to better sleep quality (Emert et al., 2017). Moreover, conscientiousness exerts a modulatory effect over neuroticism, i.e., individuals scoring high in both

neuroticism and conscientiousness would ameliorate their sleep deficiencies compared to individuals high in neuroticism but only medium or low in conscientiousness (where the negative effects of neuroticism would be maximally manifested). In general, high conscientiousness predicts better overall good health, probably because this personality trait makes these individuals more concerned about their potential illness and, therefore, enables them to seek remedy for their complaints more efficiently (Emert et al., 2017). Hence, their better sleep health habits do not necessarily imply naturally superior sleep quality per se (although maybe this is the case if shared genes linking personality and sleep were found), but simply could be the result of the prudent aspect of their personality functioning producing this beneficial effect (again, potentially creating a feedback loop between personality and sleep habits). Additionally, conscientious individuals prefer organized planning and scheduling of their lives, therefore, it is likely that they apply the same pattern to their sleep habits (i.e., being less irregular and better complying with self-adjusted bed times), providing another explanatory factor to their good sleep hygiene as a result of their personality.

On the other hand, openness was found to correlate with insomnia, but due to different causes than for neuroticism: these individuals crave intellectual stimulation (including scientific and artistic forms), and consequently, their high-strung mental activity interferes with sleep, either by postponing it until finishing their engaged activities (keeping their minds stimulated) or by causing resting problems by overflow of ideas (as mentioned, openness correlates with both intelligence and creativity) (Jauk et al., 2015, Emert et al., 2017). In fact, the term “creative insomnia” has been coined to describe a subset of these creative individuals suffering from sleep problems as they are highly devoted to their creative endeavors, e.g., Marcel Proust, Franz Kafka, Benjamin Franklin, Alexandre Dumas or Vincent van Gogh (among many other individuals of outstanding creativity) (Dechanet-Platz, 2012). A study with children found that insomnia is quite common among creative compared to control children (providing statistical validation to the claim), and even when the study did not include adults, it is very likely that the same association still holds (Healey and

Runco, 2006). Furthermore, another study found that divergent thinking (an important component of creativity) had a modest association with insomnia complaints and other related nighttime disturbances (Beaty et al., 2013). Obviously, not all creatives are necessarily insomniacs and, moreover, even if an individual is both insomniac and creative, the insomnia could be due to a different reason than their creative engagement (e.g., depression or bipolar disorder in the case of van Gogh or Kafka), nonetheless, their creative task commitment and problem-solving oriented minds could still exert a causal influence (of comparatively secondary importance). However, paradoxically, they could also try to reduce their primary causal affliction by means of their creativity (similar to psychological self-therapy) in an attempt to restore the mental instability that principally causes insomnia. Hence, the short-term bad solution (prolonged creative work and mental effort, possibly at the cost of hours of sleep) is hopefully perceived as a long-term good solution (improved mood or stamina, reduced symptoms of psychological distress or restlessness, etc., potentially treating sleep quality problems by bettering mental health), either in the arts or the sciences.

Finally, the same study also found that high trait emotional intelligence (EI, abilities related to superior emotional recognition, management, regulation, self-awareness and self-control) correlated with better sleep quality by diminished sleep disturbances (Emert et al., 2017). People in this group had less severe insomnia problems and reported (although it was only a subjective assessment) more restorative effects in their sleep. Given that EI would anti-correlate with neuroticism, the observed positive effects on sleep of the former trait were expected. Moreover, EI predicts overall good mental health and what Aristotle called *eudaimonia* (term of difficult translation, close to well-being and happiness), of which good sleep can be understood as one more facet, nonetheless, of capital importance, as it is perhaps the only facet of mental health able to interact with all other facets in bidirectional ways (Emert et al., 2017). Finally, we should not forget that not only does personality influence sleep, reciprocally, sleep also influences personality. Thus, after sleep deprivation (mainly affecting the normal

functioning of the frontal lobe, controlling or related to major aspects of personality (Ikeda et al., 2014, Schneider and Koenigs, 2017)), common among insomniacs, people become more risk-taking, impulsive, disagreeable and irrational in their decision making (as compared to their "true" self when rested, however, this true self could efface with time if the sleep problem is not treated) (Gianotti et al., 2009, Irwin, 2015). Thus, the described characteristics created by sleep deprivation resemble personality traits of low agreeableness and low conscientiousness, as if sleep loss shifted these personality dimensions (a reminder that we are our brain). We can speculate that other personality factors of the Big-5 model (Kaufman et al., 2016) could be affected by sleep loss as well, e.g., higher neuroticism (by increased stress hormone levels), higher extraversion (by disinhibition and enhanced subcortical dominance) and lower openness (by frontal deficits in cognitive flexibility and more rigid synaptic connections according to the synaptic homeostatic hypothesis) (Tononi and Cirelli, 2014). Regarding creativity, it is possible that the brain would be better able to make more remote connections (as modularity would be lower, so more "messiness" in brain networks would favor it), however, creativity is not merely about producing novel associations (divergent thinking), but also "useful" (good quality) creative products, and crucially, sleep loss would hinder the ability for its critical evaluation (as it is dependent on prefrontal cortices). More research is needed to test this proposed group of hypotheses. Moreover, inter-individual differences in the ability to cope with the effects of sleep deprivation cannot be neglected, with certain variables assuming a modulatory effect, at least for some range (e.g., higher EI subsuming a buffer against negative emotionality or higher IQ providing a counterbalance against impulsivity and risky or irrational decision making) (Emert et al., 2017).

The tight link between emotions and sleep is recognized even by the layman, yet it has not been sufficiently clarified by the neuroscientist at the mechanistic level (to be fair, the difficulty of the topic has to be acknowledged, being in the interaction of two hard topics on their own). To provide a prominent example, it is known that regular mindfulness meditation improves sleep quality and that meditators are

"more awake" (i.e., vigilant and attentive) during wakefulness, probably as a result of more efficient restorative sleep (Brewer et al., 2013, Britton et al., 2014). Hence, we propose that further studies delve deeper into this complex relationship at the neurophysiological level by applying LORETA source localization to discover statistical differences in the brain patterns of high versus low EI groups as they sleep (e.g., significantly different current spectral density, effective connectivity, etc., seizing the same tools we exposed in our project). Additionally, a longitudinal study following an insomnia group undergoing mindfulness training could unravel the putative benefits of this practice as they get better on it, with quantifiable differences assessed by LORETA statistical maps as a function of time compared to a control group (providing an approximate threshold of days to begin to produce noticeable statistical changes at the neurological level by plasticity). A better comprehension of the concrete neurobiological differences that distinguish groups of people experiencing good versus bad sleep (aided by personality research) offers tremendous therapeutic applications to improve the quality of life of millions of individuals complaining about problems for falling asleep, with negative feedback effects on the long-term in their overall psychophysical health.

## 6.3. Concluding remarks

*"Sleeping is no mean art: for its sake one must stay awake all day."*
Friedrich Nietzche[6.6]

As final remarks, we hoped through this project to have paved the path for future neuroscientific research by exploring for the first time new exciting tools for neuroscientists: eLORETA in combination with the isolated effective coherence, a very powerful tandem promising vast potential. We have shown how eLORETA straightforwardly reveals the tradeoff between delta and sigma power when statistically comparing the baseline and recovery conditions, the exponential rising of delta activity during the sleep onset transition (of even higher steepness during sleep recovery), the curvilinear progression of sigma activity and the effective connectivity breakdown of the DMN through iCOH. We finish by expressing our desire that, both for theoretical reasons and for the *eudaimonia* of society at large, we had modestly contributed to shedding light and provided new useful insights for the understanding of the enigmatic topic of sleep through this thesis.

## List of abbreviations

AASM: American academy of sleep medicine

ACC: anterior cingulate cortex

ADHD: attention deficit hyperactivity disorder

AIC: Akaike information criterion

AMI: auto-mutual information

ATP: adenosine triphosphate

BA: Brodmann area

BOLD: blood oxygen level dependent

CEN: central executive network

CMI: cross-mutual information

CNS: central nervous system

CSD: current source density

CSF: cerebrospinal fluid

DA: dimension of activation

DCM: direct causal modeling

dDTF: directed direct transfer function

DMN: default mode network

DNA: dorsal attention network

DTF: direct transfer function

DTI: diffusion tensor imaging

EC: effective connectivity

EEG: electroencephalography

eLORETA: exact low resolution electromagnetic tomography

EMG: electromyography

ERP: event-related potential

FA: fractional anisotropy

FC: functional connectivity

ffDTF: full frequency direct transfer function

fMRI: functional magnetic resonance imaging

GC: Granger causality

gPDC: generalized partial directed coherence

H: Shannon entropy

HD-EEG: high-density electroencephalography

ICA: independent component analysis

iCOH: isolated effective coherence

LDLPFC: left dorsolateral prefrontal cortex

LH: left hippocampus

LIPL: left inferior parietal lobule

LOC: loss of consciousness

LORETA: low-resolution electromagnetic tomography

MCC: middle cingulate cortex

MEG: magnetoencephalography

MI: mutual information

MNI coordinates: Montreal neurological institute coordinates

MPFC: medial prefrontal cortex

MRI: magnetic resonance imaging

MVAR modeling: multivariate autoregressive modeling

NREM sleep: non-rapid eye movement sleep

PCC: posterior cingulate cortex

PDC: partial directed coherence

PET: positron emission tomography

PGO waves: ponto-geniculo-occipital waves

PI: partial information

PMI: partial mutual information

PSD: power spectral density

PSI: phase-slope index

RDLPFC: right dorsolateral prefrontal cortex

REM sleep: rapid eye movement sleep

RH: right hippocampus

RIPL: right inferior parietal lobule

ROI: region of interest

rPDC: renormalized partial directed coherence

RSN: resting state networks

SEEG: stereotactic electroencephalography

SEM: structural equation modeling

smCSD: square-magnitude current source density

SN: saliency network

SnPM: statistical non-parametric mapping

SNR: signal-to-noise ratio

SO: sleep onset

SWA: slow wave activity

SWS: slow wave sleep

TE: transfer entropy

TMS: transcranial magnetic stimulation

# References


Adam, P., Krizkova, S., Heger, Z. *et al.* Metallothioneins in Prion- and Amyloid-Related Diseases. *Journal of Alzheimers Disease*, 2016, 51: 637-56.

Aeschbach, D. and Borbély, A. A. All-night dynamics of the human sleep EEG. *Journal of Sleep Research*, 1993, 2: 70-81.

Aeschbach, D., Dijk, D. J. and Borbély, A. A. Dynamics of EEG spindle frequency activity during extended sleep in humans: Relationship to slow-wave activity and time of day. *Brain Research*, 1997, 748: 131-36.

Aguirre, C. C. Sleep deprivation: a mind-body approach. *Current Opinion in Pulmonary Medicine*, 2016, 22: 583-88.

Akaike, H. On use of a linear model for identification of feedback systems. *Annals of the Institute of Statistical Mathematics*, 1968, 20: 425-&.

Akaike, H. A new look at statistical model identification. *Ieee Transactions on Automatic Control*, 1974, AC19: 716-23.

Alivisatos, A. P., Chun, M. Y., Church, G. M., Greenspan, R. J., Roukes, M. L. and Yuste, R. A National Network of Neurotechnology Centers for the BRAIN Initiative. *Neuron*, 2015, 88: 445-48.

Amico, E., Gomez, F., Di Perri, C. *et al.* Posterior Cingulate Cortex-Related Co-Activation Patterns: A Resting State fMRI Study in Propofol-Induced Loss of Consciousness. *Plos One*, 2014, 9

Amzica, F. and Steriade, M. The functional significance of K-complexes. *Sleep Medicine Reviews*, 2002, 6: 139-49.

Anderer, P., Klosch, G., Gruber, G. *et al.* Low-resolution brain electromagnetic tomography revealed simultaneously active frontal and parietal sleep spindle sources in the human cortex. *Neuroscience*, 2001, 103: 581-92.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R. and Buckner, R. L. Functional-Anatomic Fractionation of the Brain's Default Network. *Neuron*, 2010, 65: 550-62.

Andrillon, T., Nir, Y., Staba, R. J. *et al.* Sleep Spindles in Humans: Insights from Intracranial EEG and Unit Recordings. *Journal of Neuroscience*, 2011, 31: 17821-34.

Aserinsky, E. and Kleitman, N. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, 1953, 118: 273-74.

Ashley, R. and Tsang, K. Credible Granger-Causality inference with modest sample lengths: a cross-sample validation approach. *Econometrics 2014*, 2014, 2: 72-91.

Astolfi, L., Cincotti, F., Mattia, D. *et al.* Comparison of different cortical connectivity estimators for high-resolution EEG recordings. *Human Brain Mapping*, 2007, 28: 143-57.

Babiloni, F., Cincotti, F., Babiloni, C. *et al.* Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function. *Neuroimage*, 2005, 24: 118-31.

Baccala, L. and Sameshima, K. Using partial directed coherence to describe neuronal ensemble interactions. *J Neurosci Methods*, 1999, 94: 93-103.

Baccala, L. and Sameshima, K. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*, 2001, 84: 463-74.

Bakhshayesh, H., Fitzgibbon, S. P., Pope, K. J. and Ieee Detection of coupling with linear and nonlinear synchronization measures for EEG. In), *2nd Middle East Conference on Biomedical Engineering (MECBME)*, Doha, QATAR, 2014: 240-43.

Balachandran, D. Sleep and the Immune System: Implications for Health and Mortality. *Sleep & Safety*, 2011: 52-59.

Barbera, J. Sleep and dreaming in Greek and Roman philosophy. *Sleep Medicine*, 2008, 9: 906-10.

Barnett, L., Barrett, A. B. and Seth, A. K. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys Rev Lett*, 2009, 103: 238701.

Barrett, A., Murphy, M., Bruno, M. *et al.* Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia. *Plos One*, 2012, 7

Barrett, A. B., Barnett, L. and Seth, A. K. Multivariate Granger causality and generalized variance. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2010, 81: 041907.

Bastos, A. M. and Schoffelen, J. M. A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. *Front Syst Neurosci*, 2015, 9: 175.

Benedek, M., Schickel, R. J., Jauk, E., Fink, A. and Neubauer, A. C. Alpha power increases in right parietal cortex reflects focused internal attention. *Neuropsychologia*, 2014, 56: 393-400.

Berger, R. J. and Phillips, N. H. Energy-conservation and sleep. *Behavioural Brain Research*, 1995, 69: 65-73.

Bernasconi, C. and König, P. On the directionality of cortical interactions studied by structural analysis of electrophysiological recordings. *Biological Cybernetics*, 1999, 81: 199-210.

Berry, D. T. R. Sleep onset: Normal and abnormal processes - Ogilvie,RD, Harsh,JR. *Contemporary Psychology*, 1996, 41: 495-96.

Bersagliere, A., Pascual-Marqui, R. D., Tarokh, L. and Achermann, P. Mapping Slow Waves by EEG Topography and Source Localization: Effects of Sleep Deprivation. *Brain Topography*, 2018, 31: 257-69.

Bierwolf, C., Struve, K., Marshall, L., Born, J. and Fehm, H. L. Slow wave sleep drives inhibition of pituitary-adrenal secretion in humans. *Journal of Neuroendocrinology*, 1997, 9: 479-84.

Blinowska, K. Review of the methods of determination of directed connectivity from multichannel data. *Medical & Biological Engineering & Computing*, 2011, 49: 521-29.

Bobic, T. T., Secic, A., Zavoreo, I. *et al.* The impact of sleep deprivation on the brain. *Acta Clinica Croatica*, 2016, 55: 469-73.

Boly, M., Phillips, C., Tshibanda, L. *et al.* Intrinsic brain activity in altered states of consciousness - How conscious is the default mode of brain function? *Molecular and Biophysical Mechanisms of Arousal, Alertness, and Attention*, 2008, 1129: 119-29.

Bonhomme, V., Vanhaudenhuyse, A., Demertzi, A. *et al.* Resting-state Network-specific Breakdown of Functional Connectivity during Ketamine Alteration of Consciousness in Volunteers. *Anesthesiology*, 2016, 125: 873-88.

Borbély, A. A. A two process model of sleep regulation. *Hum Neurobiol*, 1982, 1: 195-204.

Borbély, A. A. and Achermann, P. Sleep homeostasis and models of sleep regulation. *Journal of Biological Rhythms*, 1999, 14: 557-68.

Borbély , A. A., Baumann, F., Brandeis, D., Strauch, I. and Lehmann, D. Sleep-deprivation - effect on sleep stages and EEG power-density in man. *Electroencephalography and Clinical Neurophysiology*, 1981, 51: 483-93.

Borbély, A. A., Daan, S., Wirz-Justice, A. and Deboer, T. The two-process model of sleep regulation: a reappraisal. *Journal of Sleep Research*, 2016, 25: 131-43.

Born, J., Rasch, B. and Gais, S. Sleep to remember. *Neuroscientist*, 2006, 12: 410-24.

Brazier, M. A. B. The electrical fields at the surface of the head during sleep. *Electroencephalography and Clinical Neurophysiology*, 1949, 1: 195-204.

Brewer, J. A., Garrison, K. A. and Whitfield-Gabrieli, S. What about the "self" is processed in the posterior cingulate cortex? *Frontiers in Human Neuroscience*, 2013, 7

Britton, W. B., Lindahl, J. R., Cahn, B. R., Davis, J. H. and Goldman, R. E. Awakening is not a metaphor: the effects of Buddhist meditation practices on basic wakefulness. *Advances in Meditation Research: Neuroscience and Clinical Applications*, 2014, 1307: 64-81.

Brown, R. Muramyl peptides and the functions of sleep (vol 69, pg 85, 1995). *Behavioural Brain Research*, 1997, 82: 245-45.

Bruns, A. Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? *Journal of Neuroscience Methods*, 2004, 137: 321-32.

Buckner, R. L., Andrews-Hanna, J. R. and Schacter, D. L. The brain's default network - Anatomy, function, and relevance to disease. *Year in Cognitive Neuroscience 2008*, 2008, 1124: 1-38.

Cachope, R. Functional diversity on synaptic plasticity mediated by endocannabinoids. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 2012, 367: 3242-53.

Cao, Z., Lin, C. T., Chuang, C. H. *et al.* Resting-state EEG power and coherence vary between migraine phases. *Journal of Headache and Pain*, 2016, 17

Cappuccio, F. P., D'elia, L., Strazzullo, P. and Miller, M. A. Sleep Duration and All-Cause Mortality: A Systematic Review and Meta-Analysis of Prospective Studies. *Sleep*, 2010, 33: 585-92.

Carmeli, C., Knyazeva, M. G., Innocenti, G. M. and De Feo, O. Assessment of EEG synchronization based on state-space analysis. *Neuroimage*, 2005, 25: 339-54.

Carrier, J., Land, S., Buysse, D., Kupfer, D. and Monk, T. The effects of age and gender on sleep EEG power spectral density in the middle years of life (ages 20-60 years old). *Psychophysiology*, 2001, 38: 232-42.

Cartwright, R. Normal human sleep: an overview. In: M. H. KRYGER, T. ROTH and W. C. DEMENT (Eds), *Principles and practice of sleep medicine, 3rd edition*. W. B. Saunders Company, Philadelphia, 2001: 1545-46.

Casey, S. J., Solomons, L. C., Steier, J. *et al.* Slow Wave and REM Sleep Deprivation Effects on Explicit and Implicit Memory During Sleep. *Neuropsychology*, 2016, 30: 931-45.

Cash, S. S., Halgren, E., Dehghani, N. *et al.* The Human K-Complex Represents an Isolated Cortical Down-State. *Science*, 2009, 324: 1084-87.

Cerullo, M. A. The Problem with Phi: A Critique of Integrated Information Theory. *Plos Computational Biology*, 2015, 11

Chan, H.-L., Chu, J.-H., Fung, H.-C. *et al.* Brain connectivity of patients with Alzheimer's disease by coherence and cross mutual information of electroencephalograms during photic stimulation. *Medical Engineering & Physics*, 2013, 35: 241-52.

Chaput, J. P. and Dutil, C. Lack of sleep as a contributor to obesity in adolescents: impacts on eating and activity behaviors. *International Journal of Behavioral Nutrition and Physical Activity*, 2016, 13

Chow, H. M., Horovitz, S. G., Carr, W. S. *et al.* Rhythmic alternating patterns of brain activity distinguish rapid eye movement sleep from other states of consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, 2013, 110: 10300-05.

Cirelli, C. Sleep and synaptic changes. *Current Opinion in Neurobiology*, 2013, 23: 841-46.

Cirelli, C. and Tononi, G. Sleep and Synaptic Homeostasis. *Sleep*, 2015, 38: 161-62.

Clark, I. and Landolt, H. P. Coffee, caffeine, and sleep: A systematic review of epidemiological studies and randomized controlled trials. *Sleep Medicine Reviews*, 2017, 31: 70-78.

Cohen, M. X. Comparison of different spatial transformations applied to EEG data: A case study of error processing. *International Journal of Psychophysiology*, 2015, 97: 245-57.

Connemann, B. J., Mann, K., Pascual-Marqui, R. D. and Roschke, J. Limbic activity in slow wave sleep in a healthy subject with alpha-delta sleep. *Psychiatry Research-Neuroimaging*, 2001, 107: 165-71.

Daan, S., Beersma, D. G. M. and Borbély, A. A. Timing of human sleep - recovery process gated by a circadian pacemaker. *American Journal of Physiology*, 1984, 246: R161-R78.

Dalal, S. S., Hamame, C. M., Eichenlaub, J. B. and Jerbi, K. Intrinsic Coupling between Gamma Oscillations, Neuronal Discharges, and Slow Cortical Oscillations during Human Slow-Wave Sleep. *Journal of Neuroscience*, 2010, 30: 14285-87.

Dang-Vu, T. T., Schabus, M., Desseilles, M. *et al.* Spontaneous neural activity during human slow wave sleep. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105: 15160-65.

Davis, H., Davis, P. A., Loomis, A. L., Harvey, E. N. and Hobart, G. Human brain potentials during the onset of sleep. *Journal of Neurophysiology*, 1938, 1: 24-38.

De Gennaro, L. and Ferrara, M. Sleep spindles: an overview. *Sleep Medicine Reviews*, 2003, 7: 423-40.

De Gennaro, L., Ferrara, M. and Bertini, M. Effect of slow-wave sleep deprivation on topographical distribution of spindles. *Behavioural Brain Research*, 2000a, 116: 55-59.

De Gennaro, L., Ferrara, M. and Bertini, M. The boundary between wakefulness and steep: Quantitative electroencephalographic changes during the sleep onset period. *Neuroscience*, 2001a, 107: 1-11.

De Gennaro, L., Ferrara, M., Curcio, G. and Cristiani, R. Antero-posterior EEG changes during the wakefulness-sleep transition. *Clinical Neurophysiology*, 2001b, 112: 1901-11.

De Gennaro, L., Ferrara, M., Ferlazzo, F. and Bertini, M. Slow eye movements and EEG power spectra during wake-sleep transition. *Clinical Neurophysiology*, 2000b, 111: 2107-15.

De Gennaro, L., Ferrara, M., Vecchio, F., Curcio, G. and Bertini, M. An electroencephalographic fingerprint of human sleep. *Neuroimage*, 2005a, 26: 114-22.

De Gennaro, L., Vecchio, F., Ferrara, M., Curcio, G., Rossini, P. M. and Babiloni, C. Changes in fronto-posterior functional coupling at sleep onset in humans. *Journal of Sleep Research*, 2004, 13: 209-17.

De Gennaro, L., Vecchio, F., Ferrara, M., Curcio, G., Rossini, P. M. and Babiloni, C. Antero-posterior functional coupling at sleep onset: changes as a function of increased sleep pressure. *Brain Research Bulletin*, 2005b, 65: 133-40.

De Munck, J. C., Goncalves, S. I., Huijboom, L. *et al.* The hemodynamic response of the alpha rhythm: An EEG/fMRI study. *Neuroimage*, 2007, 35: 1142-51.

Del Felice, A., Arcaro, C., Storti, S. F., Fiaschi, A. and Manganotti, P. Electrical Source Imaging of Sleep Spindles. *Clinical Eeg and Neuroscience*, 2014, 45: 184-92.

Dijk, D. J., Brunner, D. P. and Borbely, A. A. Time course of EEG power-density during long sleep in humans. *American Journal of Physiology*, 1990, 258: R650-R61.

Dipasquale, O. and Cercignani, M. Network functional connectivity and whole-brain functional connectomics to investigate cognitive decline in neurodegenerative conditions. *Functional Neurology*, 2016, 31: 191-203.

Dissel, S., Seugnet, L., Thimgan, M. S. *et al.* Differential activation of immune factors in neurons and glia contribute to individual differences in resilience/vulnerability to sleep disruption. *Brain Behavior and Immunity*, 2015, 47: 75-85.

Djordjevic, Z., Jovanovic, A. and Perovic, A. Brain connectivity measure -the Direct Transfer Function- advantages and weak points. *Intelligent Systems and Informatics (SISY), 2012 IEEE 10th Jubilee International Symposium*, 2012: 93-97.

Dolezal, B. A., Neufeld, E. V., Boland, D. M., Martin, J. L. and Cooper, C. B. Interrelationship between Sleep and Exercise: A Systematic Review. *Adv Prev Med*, 2017, 2017

Douw, L., Nieboer, D., Van Dijk, B. W., Stam, C. J. and Twisk, J. W. R. A Healthy Brain in a Healthy Body: Brain Network Correlates of Physical and Mental Fitness. *Plos One*, 2014, 9

Eckert, M. and Tatsuno, M. Overview of Neural Activity in the Awake and Sleeping Hippocampus. *Analysis and Modeling of Coordinated Multi-Neuronal Activity*, 2015, 12: 65-79.

Edlow, B. L. and Lammers, G. J. Bringing posttraumatic sleep-wake disorders out of the dark. *Neurology*, 2016, 86: 1934-35.

Erla, S., Faes, L., Tranquillini, E., Orrico, D. and Nollo, G. Multivariate autoregressive model with instantaneous effects to improve brain connectivity estimation. *International Journal of Bioelectromagnetism*, 2009, 11: 74-79.

Escudero, J., Hornero, R. and Abasolo, D. Interpretation of the auto-mutual information rate of decrease in the context of biomedical signal analysis. Application to electroencephalogram recordings. *Physiological Measurement*, 2009, 30: 187-99.

Esser, S. K., Hill, S. and Tononi, G. Breakdown of Effective Connectivity During Slow Wave Sleep: Investigating the Mechanism Underlying a Cortical Gate Using Large-Scale Modeling. *Journal of Neurophysiology*, 2009, 102: 2096-111.

Esslen, M., Metzler, S., Pascual-Marqui, R. and Jancke, L. Pre-reflective and reflective self-reference: A spatiotemporal EEG analysis. *Neuroimage*, 2008a, 42: 437-49.

Esslen, M., Metzler, S., Pascual-Marqui, R. and Jäncke, L. Pre-reflective and reflective self-reference: A spatiotemporal EEG analysis. *Neuroimage*, 2008b, 42: 437-49.

Everson, C. A., Bergmann, B. M. and Rechtschaffen, A. Sleep-deprivation in the rat: III. Total sleep deprivation. *Sleep*, 1989, 12: 13-21.

Ewald, A., Avarvand, F. S. and Nolte, G. Identifying causal networks of neuronal sources from EEG/MEG data with the phase slope index: a simulation study. *Biomedical Engineering-Biomedizinische Technik*, 2013, 58: 165-78.

Faes, L., Erla, S. and Nollo, G. Measuring connectivity in linear multivariate processes: definitions, interpretation, and practical analysis. *Computational and Mathematical Methods in Medicine*, 2012

Faes, L., Erla, S., Tranquillini, E., Orrico, D. and Nollo, G. An identifiable model to assess frequency-domain Granger causality in the presence of significant instantaneous interactions. *Conf Proc IEEE Eng Med Biol Soc*, 2010, 2010: 1699-702.

Faes, L., Marinazzo, D., Jurysta, F. and Nollo, G. Granger causality analysis of sleep brain-heart interactions. *2014 8th Conference of the European Study Group on Cardiovascular Oscillations (Esgco)*, 2014: 5-6.

Faes, L. and Nollo, G. Measuring frequency domain Granger causality for multiple blocks of interacting time series. *Biological Cybernetics*, 2013, 107: 217-32.

Faes, L., Porta, A. and Nollo, G. Surrogate data approaches to assess the significance of directed coherence: application to EEG activity propagation. *Conf Proc IEEE Eng Med Biol Soc*, 2009, 2009: 6280-3.

Farthouat, J. and Peigneux, P. *Memory Reactivation in Humans (Imaging Studies)*. 2015 (

Ferrara, M. and De Gennaro, L. Going Local: Insights from EEG and Stereo-EEG Studies of the Human Sleep-Wake Cycle. *Current Topics in Medicinal Chemistry*, 2011, 11: 2423-37.

Finelli, L. A., Achermann, P. and Borbély, A. A. Individual "fingerprints" in human sleep EEG topography. *Neuropsychopharmacology*, 2001a, 25: S57-S62.

Finelli, L. A., Baumann, H., Borbély, A. A. and Achermann, P. Dual electroencephalogram markers of human sleep homeostasis: Correlation between theta activity in waking and slow-wave activity in sleep. *Neuroscience*, 2000, 101: 523-29.

Finelli, L. A., Borbély, A. A. and Achermann, P. Functional topography of the human nonREM sleep electroencephalogram. *European Journal of Neuroscience*, 2001b, 13: 2282-90.

Fonseca, A., Boboeva, V., Brederoo, S. and Baggio, G. Disrupting morphosyntactic and lexical semantic processing has opposite effects on the sample entropy of neural signals. *Brain Research*, 2015, 1604: 1-14.

Forget, D., Morin, C. M. and Bastien, C. H. The Role of the Spontaneous and Evoked K-Complex in Good-Sleeper Controls and in Individuals with Insomnia. *Sleep*, 2011, 34: 1251-60.

Franzen, J. D., Heinrichs-Graham, E., White, M. L., Wetzel, M. W., Knott, N. L. and Wilson, T. W. Atypical coupling between posterior regions of the default mode network in attention-deficit/hyperactivity disorder: a pharmaco-magnetoencephalography study. *Journal of Psychiatry & Neuroscience*, 2013, 38: 333-40.

Frenzel, S. and Pompe, B. Partial mutual information for coupling analysis of multivariate time series. *Physical Review Letters*, 2007, 99

Friston, K. J. Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 1994, 2: 56-78.

Friston, K. J. Functional and effective connectivity: a review. *Brain connectivity*, 2011, 1: 13-36.

Friston, K. J., Bastos, A. M., Oswal, A., Van Wijk, B., Richter, C. and Litvak, V. Granger causality revisited. *Neuroimage*, 2014, 101: 796-808.

Gais, S. and Born, J. Declarative memory consolidation: Mechanisms acting during human sleep. *Learning & Memory*, 2004, 11: 679-85.

Garcia-Rill, E., Wallace, T. and Good, C. Neuropharmacology of sleep and wakefulness. *Sleep: a Comprehensive Handbook*, 2006: 63-71.

Geweke, J. Measurement of linear-dependence and feedback between multiple time-series. *Journal of the American Statistical Association*, 1982, 77: 304-13.

Gianotti, L. R., Knoch, D., Faber, P. L. *et al.* Tonic Activity Level in the Right Prefrontal Cortex Predicts Individuals' Risk Taking. *Psychological Science*, 2009, 20: 33-38.

Glasser, M. F., Smith, S. M., Marcus, D. S. *et al.* The Human Connectome Project's neuroimaging approach. *Nature Neuroscience*, 2016, 19: 1175-87.

Goder, R., Seeck-Hirschner, M., Stingele, K. *et al.* Sleep and cognition at baseline and the effects of REM sleep diminution after 1 week of antidepressive treatment in patients with depression. *Journal of Sleep Research*, 2011, 20: 544-51.

Goldenberg, D. and Galvan, A. The use of functional and effective connectivity techniques to understand the developing brain. *Developmental Cognitive Neuroscience*, 2015, 12: 155-64.

Gomez, C. M., Marco, J. and Grau, C. Preparatory visuo-motor cortical network of the contingent negative variation estimated by current density. *Neuroimage*, 2003, 20: 216-24.

Granger, C. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 1969, 37: 414-&.

Grech, R., Cassar, T., Muscat, J. *et al.* Review on solving the inverse problem in EEG source analysis. *Journal of Neuroengineering and Rehabilitation*, 2008, 5

He, Y. and Evans, A. Graph theoretical modeling of brain connectivity. *Current Opinion in Neurology*, 2010, 23: 341-50.

Heine, L., Soddu, A., Gomez, F. *et al.* Resting state networks and consciousness Alterations of multiple resting state network connectivity in physiological, pharmacological, and pathological consciousness states. *Frontiers in Psychology*, 2012, 3

Herbet, G., Lafargue, G., De Champfleur, N. M. *et al.* Disrupting posterior cingulate connectivity disconnects consciousness from the external environment. *Neuropsychologia*, 2014, 56: 239-44.

Herbet, G., Lafargue, G. and Duffau, H. The dorsal cingulate cortex as a critical gateway in the network supporting conscious awareness. *Brain*, 2016, 139

Hlavackova-Schindler, K., Palus, M., Vejmelka, M. and Bhattacharya, J. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports-Review Section of Physics Letters*, 2007, 441: 1-46.

Hobson, J. A. REM sleep and dreaming: towards a theory of protoconsciousness. *Nature Reviews Neuroscience*, 2009, 10: 803-U62.

Hobson, J. A. and Friston, K. J. Waking and dreaming consciousness: Neurobiological and functional considerations. *Progress in Neurobiology*, 2012, 98: 82-98.

Hobson, J. A. and Mccarley, R. W. The Brain as a Dream State Generator: An Activation-Synthesis Hypothesis of the Dream Process. *Essential Sources in the Scientific Study of Consciousness*, 2003: 937-57.

Hodor, A., Palchykova, S., Baracchi, F., Horvath, T. and Bassetti, C. L. Sleep promotion with baclofen improves functional recovery and promotes neuroplasticity after stroke in rats. *Journal of Neurology*, 2014, 261: S212-S13.

Hori, T. Spatiotemporal changes of EEG activity during waking-sleeping transition period. *International Journal of Neuroscience*, 1985, 27: 101-14.

Horovitz, S. G., Braun, A. R., Carr, W. S. *et al.* Decoupling of the brain's default mode network during deep sleep. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106: 11376-81.

Houck, J. M., Cetin, M. S., Mayer, A. R. *et al.* Magnetoencephalographic and functional MRI connectomics in schizophrenia via intra- and inter-network connectivity. *Neuroimage*, 2017, 145: 96-106.

Huang, C. S., Pal, N. R., Chuang, C. H. and Lin, C. T. Identifying changes in EEG information transfer during drowsy driving by transfer entropy. *Frontiers in Human Neuroscience*, 2015, 9

Huang, D. F., Ren, A. F., Shang, J. *et al.* Combining Partial Directed Coherence and Graph Theory to Analyse Effective Brain Networks of Different Mental Tasks. *Frontiers in Human Neuroscience*, 2016, 10

Hutchison, I. C. and Rathore, S. The role of REM sleep theta activity in emotional memory. *Frontiers in Psychology*, 2015, 6

Iber, C., Ancoli-Israel, S., Chesson, A. L. and Quan, S. F. The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications. *Westchester,Illinois, USA: American Academy of Sleep Medicine*, 2007a

Iber, C., Ancoli-Israel, S., Chesson, A. L. and Quan, S. F. The New Sleep Scoring Manual-The Evidence Behind The Rules. *Journal of Clinical Sleep Medicine*, 2007b, 3: 107-07.

Ikeda, H., Ikeda, E., Shiozaki, K. and Hirayasu, Y. Association of the five-factor personality model with prefrontal activation during frontal lobe task performance using two-channel near-infrared spectroscopy. *Psychiatry and Clinical Neurosciences*, 2014, 68: 752-58.

Ikeda, S., Mizuno-Matsumoto, Y., Canuet, L. *et al.* Emotion Regulation of Neuroticism: Emotional Information Processing Related to Psychosomatic State Evaluated by Electroencephalography and Exact Low-Resolution Brain Electromagnetic Tomography. *Neuropsychobiology*, 2015, 71: 34-41.

Irwin, M. R. Why Sleep Is Important for Health: A Psychoneuroimmunology Perspective. *Annual Review of Psychology, Vol 66*, 2015, 66: 143-72.

Irwin, M. R., Levin, M. J., Carrillo, C. *et al.* Major depressive disorder and immunity to varicella-zoster virus in the elderly. *Brain Behavior and Immunity*, 2011, 25: 759-66.

Irwin, M. R. and Opp, M. R. Sleep Health: Reciprocal Regulation of Sleep and Innate Immunity. *Neuropsychopharmacology*, 2017, 42: 129-55.

Jatoi, M. A., Kamel, N., Malik, A. S. and Faye, I. EEG based brain source localization comparison of sLORETA and eLORETA. *Australasian Physical & Engineering Sciences in Medicine*, 2014, 37: 713-21.

Jauk, E., Neubauer, A. C., Dunst, B., Fink, A. and Benedek, M. Gray matter correlates of creative potential: A latent variable voxel-based morphometry study. *Neuroimage*, 2015, 111: 312-20.

Jeong, J., Gore, J. and Peterson, B. Mutual information analysis of the EEG in patients with Alzheimer's disease. *Clinical Neurophysiology*, 2001, 112: 827-35.

Jessen, N. A., Munk, A. S. F., Lundgaard, I. and Nedergaard, M. The Glymphatic System: A Beginner's Guide. *Neurochemical Research*, 2015, 40: 2583-99.

Jihene, B., Wassim, Z. and Ahmed, B. Effect of Introducing time component on the EEG inverse problem. *2013 International Conference on Individual and Collective Behaviors in Robotics (Icbr)*, 2013: 75-79.

Jimenovaldes, A. and Jimenobulnes, N. Creativity and the hypnagogic state. *Folia Humanistica*, 1993, 31: 129-47.

Jones, B. E. Neuroscience: What Are Cortical Neurons Doing during Sleep? *Current Biology*, 2016, 26: R1147-R50.

Joshi, S. S., Lesser, T. J., Olsen, J. W. and O'hara, B. F. The importance of temperature and thermoregulation for optimal human sleep. *Energy and Buildings*, 2016, 131: 153-57.

Julitta, B., Vallverdu, M., Melia, U. S. *et al.* Auto-mutual information function of the EEG as a measure of depth of anesthesia. *Conf Proc IEEE Eng Med Biol Soc*, 2011, 2011: 2574-7.

Jung, C. G. Psychological Types. *Psychological Types*, 1921: 1-548.

Jung, C. M., Melanson, E. L., Frydendall, E. J., Perreault, L., Eckel, R. H. and Wright, K. P. Energy expenditure during sleep, sleep deprivation and sleep following sleep deprivation in adult humans. *Journal of Physiology-London*, 2011, 589: 235-44.

Jäncke, L. Music drives brain plasticity. *International Journal of Psychophysiology*, 2016, 108: 46-46.

Kalinchuk, A. V., Mccarley, R. W., Porkka-Heiskanen, T. and Basheer, R. Sleep Deprivation Triggers Inducible Nitric Oxide-Dependent Nitric Oxide Production in Wake-Active Basal Forebrain Neurons. *Journal of Neuroscience*, 2010, 30: 13254-64.

Kaminski, M. and Blinowska, K. A new method of the description of the information-flow in the brain structures. *Biological Cybernetics*, 1991, 65: 203-10.

Kaminski, M. and Blinowska, K. Directed Transfer Function is not influenced by volume conduction-inexpedient pre-processing should be avoided. *Frontiers in Computational Neuroscience*, 2014, 8

Kaminski, M., Blinowska, K. and Szelenberger, W. Topographic analysis of coherence and propagation of EEG activity during sleep and wakefulness. *Electroencephalography and Clinical Neurophysiology*, 1997, 102: 216-27.

Kaminski, M., Ding, M., Truccolo, W. and Bressler, S. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics*, 2001, 85: 145-57.

Karlsson, B., Hassan, M. and Marque, C. Windowed multivariate autoregressive model improving classification of labor vs. pregnancy contractions. *Conf Proc IEEE Eng Med Biol Soc*, 2013, 2013: 7444-7.

Kavanau, J. L. Origin and evolution of sleep: Roles of vision and endothermy. *Brain Research Bulletin*, 1997, 42: 245-64.

Khadem, A. and Hossein-Zadeh, G. A. Estimation of direct nonlinear effective connectivity using information theory and multilayer perceptron. *J Neurosci Methods*, 2014, 229: 53-67.

Kim, H., Hudetz, A. G., Lee, J., Mashour, G. A., Lee, U. and Re, C. S. G. Estimating the Integrated Information Measure Phi from High-Density Electroencephalography during States of Consciousness in Humans. *Frontiers in Human Neuroscience*, 2018, 12

Kim, U., Bal, T. and Mccormick, D. A. Spindle waves are propagating synchronized oscillations in the ferret LGND in-vitro. *Journal of Neurophysiology*, 1995, 74: 1301-23.

Kinney, J. B. and Atwal, G. S. Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111: 3354-59.

Klemm, W. R. Why does rem sleep occur? A wake-up hypothesis. *Front Syst Neurosci*, 2011, 5: 73.

Klimesch, W., Sauseng, P. and Hanslmayr, S. EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Research Reviews*, 2007, 53: 63-88.

Kocsis, B., Bragin, A. and Buzsaki, G. Interdependence of multiple theta generators in the hippocampus: a partial coherence analysis. *Journal of Neuroscience*, 1999, 19: 6200-12.

Kortner, G. and Geiser, F. The temporal organization of daily torpor and hibernation: Circadian and circannual rhythms. *Chronobiology International*, 2000, 17: 103-28.

Korzeniewska, A., Crainiceanu, C., Kus, R., Franaszczuk, P. and Crone, N. Dynamics of Event-Related Causality in brain electrical activity. *Human Brain Mapping*, 2008, 29: 1170-92.

Korzeniewska, A., Mańczak, M., Kamiński, M., Blinowska, K. J. and Kasicki, S. Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method. *J Neurosci Methods*, 2003, 125: 195-207.

Kralemann, B., Pikovsky, A. and Rosenblum, M. Reconstructing effective phase connectivity of oscillator networks from observations. *New Journal of Physics*, 2014, 16

Krishnaswamy, P., Obregon-Henao, G., Ahveninen, J. *et al.* Sparsity enables estimation of both subcortical and cortical activity from MEG and EEG. *Proceedings of the*

*National Academy of Sciences of the United States of America*, 2017, 114: E10465-E74.

Kristan, W. B. Functional Connectomics: How Maggots Make Up Their Minds. *Current Biology*, 2017, 27: R38-R41.

Krueger, J. M., Frank, M. G., Wisor, J. P. and Roy, S. Sleep function: Toward elucidating an enigma. *Sleep Medicine Reviews*, 2016, 28: 46-54.

Kuhn, S., Ritter, S. M., Muller, B. C. N., Van Baaren, R. B., Brass, M. and Dijksterhuis, A. The Importance of the Default Mode Network in Creativity-A Structural MRI Study. *Journal of Creative Behavior*, 2014, 48: 152-63.

Kunhimangalam, R., Joseph, P. K. and Sujith, O. K. Nonlinear analysis of EEG signals: Surrogate data analysis. *Irbm*, 2008, 29: 239-44.

Kuo, T. B. J., Chen, C. Y., Hsu, Y. C. and Yang, C. C. H. EEG beta power and heart rate variability describe the association between cortical and autonomic arousals across sleep. *Autonomic Neuroscience-Basic & Clinical*, 2016, 194: 32-37.

Ladenbauer, J., Kulzow, N., Passmann, S. *et al.* Brain stimulation during an afternoon nap boosts slow oscillatory activity and memory consolidation in older adults. *Neuroimage*, 2016, 142: 301-13.

Landolt, H. P. Sleep homeostasis: A role for adenosine in humans? *Biochemical Pharmacology*, 2008, 75: 2070-79.

Landolt, H. P. "No Thanks, Coffee Keeps Me Awake": Individual Caffeine Sensitivity Depends on ADORA2A Genotype. *Sleep*, 2012, 35: 899-900.

Landolt, H. P. Caffeine, the circadian clock, and sleep. *Science*, 2015, 349: 1289-89.

Landolt, H. P., Dijk, D. J., Gaus, S. E. and Borbély, A. A. Caffeine reduces low-frequency delta-activity in the human sleep EEG. *Neuropsychopharmacology*, 1995, 12: 229-38.

Lange, T., Dimitrov, S., Bollinger, T., Diekelmann, S. and Born, J. Sleep after Vaccination Boosts Immunological Memory. *Journal of Immunology*, 2011, 187: 283-90.

Langer, N., Pedroni, A., Gianotti, L. R., Hanggi, J., Knoch, D. and Jancke, L. Functional brain network efficiency predicts intelligence. *Human Brain Mapping*, 2012, 33: 1393-406.

Lantz, G., Michel, C. M., Pascual-Marqui, R. D. *et al.* Extracranial localization of intracranial interictal epileptiform activity using LORETA (low resolution electromagnetic tomography). *Electroencephalography and Clinical Neurophysiology*, 1997, 102: 414-22.

Larson-Prior, L. J., Power, J. D., Vincent, J. L. *et al.* Modulation of the brain's functional network architecture in the transition from wake to sleep. *Slow Brain Oscillations of Sleep, Resting State and Vigilance*, 2011, 193: 277-94.

Lavenex, P. and Amaral, D. G. Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus*, 2000, 10: 420-30.

Le Van Quyen, M., Foucher, J., Lachaux, J. *et al.* Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. *Journal of Neuroscience Methods*, 2001, 111: 83-98.

Lee, J., Nemati, S., Silva, I., Edwards, B., Butler, J. and Malhotra, A. Transfer Entropy Estimation and Directional Coupling Change Detection in Biomedical Time Series. *Biomedical Engineering Online*, 2012, 11

Leech, R. and Sharp, D. J. The role of the posterior cingulate cortex in cognition and disease. *Brain*, 2014, 137: 12-32.

Lesku, J. A., Roth, T. C., Rattenborg, N. C., Amlaner, C. J. and Lima, S. L. History and future of comparative analyses in sleep research. *Neuroscience and Biobehavioral Reviews*, 2009, 33: 1024-36.

Lesku, J. A., Vyssotski, A. L., Martinez-Gonzalez, D., Wilzeck, C. and Rattenborg, N. C. Local sleep homeostasis in the avian brain: convergence of sleep function in mammals and birds? *Proceedings of the Royal Society B-Biological Sciences*, 2011, 278: 2419-28.

Leszczynski, M. How does hippocampus contribute to working memory processing? *Frontiers in Human Neuroscience*, 2011, 5

Li, X. F., Zhou, Y. X. and Zhang, L. Newborns' sleep-wake cycle development on amplitude integrated electroencephalography. *World Journal of Pediatrics*, 2016, 12: 327-34.

Liang, H., Ding, M., Nakamura, R. and Bressler, S. Causal influences in primate cerebral cortex during visual pattern discrimination. *Neuroreport*, 2000, 11: 2875-80.

Lim, A. S., Lozano, A. M., Moro, E. *et al.* Characterization of REM-sleep associated ponto-geniculo-occipital waves in the human pons. *Sleep*, 2007, 30: 823-27.

Lin, F. H., Chu, Y. H., Hsu, Y. C. *et al.* Significant feed-forward connectivity revealed by high frequency components of BOLD fMRI signals. *Neuroimage*, 2015, 121: 69-77.

Liu, Q., Balsters, J. H., Baechinger, M., Van Der Groen, O., Wenderoth, N. and Mantini, D. Estimating a neutral reference for electroencephalographic recordings: the importance of using a high-density montage and a realistic head model. *Journal of neural engineering*, 2015, 12: 056012-12.

Liu, Y. and Aviyente, S. Quantification of effective connectivity in the brain using a measure of Directed Information. *Computational and Mathematical Methods in Medicine*, 2012

Magnin, M., Rey, M., Bastuji, H., Guillemant, P., Mauguiere, F. and Garcia-Larrea, L. Thalamic deactivation at sleep onset precedes that of the cerebral cortex in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 2010, 107: 3829-33.

Magosso, E., Ursino, M., Provini, F. and Montagna, P. Wavelet analysis of electroencephalographic and electro-oculographic changes during the sleep onset period. *2007 Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vols 1-16*, 2007: 4006-10.

Maret, S., Dorsaz, S., Gurcel, L. *et al.* Homer1a is a core brain molecular correlate of sleep loss. *Sleep*, 2008, 31: A360-A60.

Markram, H. The human brain project. *Scientific American*, 2012, 306: 50-55.

Marzano, C., Moroni, F., Gorgoni, M., Nobili, L., Ferrara, M. and De Gennaro, L. How we fall asleep: regional and temporal differences in electroencephalographic synchronization at sleep onset. *Sleep Medicine*, 2013, 14: 1112-22.

Mashour, G. Sleep, Anesthesia, and Consciousness. *Sleep*, 2011, 34: 247-48.

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H. and Tononi, G. Breakdown of cortical effective connectivity during sleep. *Science*, 2005, 309: 2228-32.

May, R. J., Maier, H. R., Dandy, G. C. and Fernando, T. M. K. G. Non-linear variable selection for artificial neural networks using partial mutual information. *Environmental Modelling & Software*, 2008, 23: 1312-26.

Mednick, S. C., Mcdevitt, E. A., Walsh, J. K. *et al.* The Critical Role of Sleep Spindles in Hippocampal-Dependent Memory: A Pharmacology Study. *Journal of Neuroscience*, 2013, 33: 4494-504.

Meerlo, P., Mistiberger, R. E., Jacobs, B. L., Heller, H. C. and Mcginty, D. New neurons in the adult brain: The role of sleep and consequences of sleep loss. *Sleep Medicine Reviews*, 2009, 13: 187-94.

Meisel, C., Olbrich, E., Shriki, O. and Achermann, P. Fading Signatures of Critical Brain Dynamics during Sustained Wakefulness in Humans. *Journal of Neuroscience*, 2013, 33: 17363-72.

Melia, U., Guaita, M., Vallverdu, M. *et al.* Mutual information measures applied to EEG signals for sleepiness characterization. *Medical Engineering & Physics*, 2015, 37: 297-308.

Mendelsohn, A. R. and Larrick, J. W. Sleep Facilitates Clearance of Metabolites from the Brain: Glymphatic Function in Aging and Neurodegenerative Diseases. *Rejuvenation Research*, 2013, 16: 518-23.

Merica, H. and Gaillard, J. M. The EEG of the sleep onset period in insomnia - a discriminant analysis. *Physiology & Behavior*, 1992, 52: 199-204.

Murphy, M., Bruno, M. A., Riedner, B. A. *et al.* Propofol Anesthesia and Sleep: A High-Density EEG Study. *Sleep*, 2011a, 34: 283-U183.

Murphy, M., Huber, R., Esser, S. *et al.* The Cortical Topography of Local Sleep. *Current Topics in Medicinal Chemistry*, 2011b, 11: 2438-46.

Na, S. H., Jin, S. H. and Kim, S. Y. The effects of total sleep deprivation on brain functional organization: Mutual information analysis of waking human EEG. *International Journal of Psychophysiology*, 2006, 62: 238-42.

Nichols, A. L. A., Eichler, T., Latham, R. and Zimmer, M. A global brain state underlies C. elegans sleep behavior. *Science*, 2017, 356: 1247-+.

Nichols, T. E. and Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp*, 2002, 15: 1-25.

Nir, Y., Staba, R. J., Andrillon, T. *et al.* Regional Slow Waves and Spindles in Human Sleep. *Neuron*, 2011, 70: 153-69.

Nir, Y. and Tononi, G. Dreaming and the brain: from phenomenology to neurophysiology. *Trends in Cognitive Sciences*, 2010, 14: 88-100.

Ogilvie, R. D. The process of falling asleep. *Sleep Medicine Reviews*, 2001, 5: 247-70.

Olbrich, E. and Achermann, P. Analysis of oscillatory patterns in the human sleep EEG using a novel detection algorithm. *Journal of Sleep Research*, 2005, 14: 337-46.

Olbrich, E., Landolt, H. P. and Achermann, P. Effect of prolonged wakefulness on electroencephalographic oscillatory activity during sleep. *Journal of Sleep Research*, 2014, 23: 253-60.

Oliviero, A., Tecchio, F., Zappasodi, F. *et al.* Brain sensorimotor hand area functionality in acute stroke: insights from magnetoencephalography. *Neuroimage*, 2004, 23: 542-50.

Ong, J. L., Kong, D. Y., Chia, T. T. Y., Tandi, J., Yeo, B. T. T. and Chee, M. W. L. Co-activated yet disconnected-Neural correlates of eye closures when trying to stay awake. *Neuroimage*, 2015, 118: 553-62.

Osborne, D. M., Pearson-Leary, J. and Mcnay, E. C. The neuroenergetics of stress hormones in the hippocampus and implications for memory. *Frontiers in Neuroscience*, 2015, 9

Paavolainen, L., Acar, E., Tuna, U. *et al.* Compensation of Missing Wedge Effects with Sequential Statistical Reconstruction in Electron Tomography. *Plos One*, 2014, 9

Palus, M. Detecting phase synchronization in noisy systems. *Physics Letters a*, 1997, 235: 341-51.

Park, D. H., Ha, J. H., Ryu, S. H., Yu, J. and Shin, C. J. Three-Dimensional Electroencephalographic Changes on Low-Resolution Brain Electromagnetic Tomography (LORETA) During the Sleep Onset Period. *Clinical Eeg and Neuroscience*, 2015, 46: 340-46.

Park, H. J. and Friston, K. J. Structural and Functional Brain Networks: From Connections to Cognition. *Science*, 2013, 342: 579-+.

Parmeggiani, P. L. REM sleep related increase in brain temperature: A physiologic problem. *Archives Italiennes De Biologie*, 2007, 145: 13-21.

Parrino, L., Ferri, R., Zucconi, M. and Fanfulla, F. Commentary from the Italian Association of Sleep Medicine on the AASM manual for the scoring of sleep and associated events: For debate and discussion. *Sleep Medicine*, 2009, 10: 799-808.

Pascual-Marqui, R. Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods and Findings in Experimental and Clinical Pharmacology*, 2002, 24: 5-12.

Pascual-Marqui, R., Biscay, R., Bosch-Bayard, J. *et al.* Assessing direct paths of intracortical causal information flow of oscillatory activity with the isolated effective coherence (iCoh). *Frontiers in Human Neuroscience*, 2014a, 8

Pascual-Marqui, R. D., Esslen, M., Kochi, K. and Lehmann, D. Functional imaging with low-resolution brain electromagnetic tomography (LORETA): A review. *Methods and Findings in Experimental and Clinical Pharmacology*, 2002, 24: 91-95.

Pascual-Marqui, R. D., Lehmann, D., Koukkou, M. *et al.* Assessing interactions in the brain with exact low-resolution electromagnetic tomography. *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, 2011, 369: 3768-84.

Pascual-Marqui, R. D., Michel, C. M. and Lehmann, D. Low-resolution electromagnetic tomography - a new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, 1994, 18: 49-65.

Pascual-Marqui, R. D., Rolando, J. B., Bosch-Bayard, J. *et al.* Advances in EEG methods applied to intra-cortical connectivity inference and to functional imaging: Examples in psychiatry research. *International Journal of Psychophysiology*, 2014b, 94: 121-21.

Penny, W. D., Litvak, V., Fuentemilla, L., Duzel, E. and Friston, K. J. Dynamic Causal Models for phase coupling. *Journal of Neuroscience Methods*, 2009, 183: 19-30.

Philiastides, M. and Sajda, P. Causal influences in the human brain during face discrimination: A short-window directed transfer function approach. *Ieee Transactions on Biomedical Engineering*, 2006, 53: 2602-05.

Piano, C., Mazzucchi, E., Bentivoglio, A. R. *et al.* Wake and Sleep EEG in Patients With Huntington Disease: An eLORETA Study and Review of the Literature. *Clinical Eeg and Neuroscience*, 2017, 48: 60-71.

Picchioni, D., Fukunaga, M., Carr, W. S. *et al.* fMRI differences between early and late stage-1 sleep. *Neuroscience Letters*, 2008, 441: 81-85.

Pires, C. A. L. and Perdigao, R. A. P. Minimum Mutual Information and Non-Gaussianity Through the Maximum Entropy Method: Theory and Properties. *Entropy*, 2012, 14: 1103-26.

Pizzagalli, D., Lehmann, D., Koenig, T., Regard, M. and Pascual-Marqui, R. D. Face-elicited ERPs and affective attitude: brain electric microstate and tomography analyses. *Clinical Neurophysiology*, 2000, 111: 521-31.

Plog, B. A., Dashnaw, M. L., Hitomi, E. *et al.* Biomarkers of Traumatic Injury Are Transported from Brain to Blood via the Glymphatic System. *Journal of Neuroscience*, 2015, 35: 518-26.

Porkka-Heiskanen, T. Sleep homeostasis. *Current Opinion in Neurobiology*, 2013, 23: 799-805.

Prerau, M. J., Hartnack, K. E., Obregon-Henao, G. *et al.* Tracking the Sleep Onset Process: An Empirical Model of Behavioral and Physiological Dynamics. *Plos Computational Biology*, 2014, 10

Ramanand, P., Bruce, M. and Bruce, E. Mutual information analysis of EEG signals indicates age-related changes in cortical interdependence during sleep in middle-aged versus elderly women. *Journal of Clinical Neurophysiology*, 2010a, 27: 274-84.

Ramanand, P., Bruce, M. and Bruce, E. Transient decoupling of cortical EEGs following arousals during NREM sleep in middle-aged and elderly women. *International Journal of Psychophysiology*, 2010b, 77: 71-82.

Rasch, B. and Born, J. About sleep's role in memory. *Physiological Reviews*, 2013, 93: 681-766.

Rattenborg, N. C., Amlaner, C. J. and Lima, S. L. Behavioral, neurophysiological and evolutionary perspectives on unihemispheric sleep. *Neuroscience and Biobehavioral Reviews*, 2000, 24: 817-42.

Rattenborg, N. C., Martinez-Gonzalez, D. and Lesku, J. A. Avian sleep homeostasis: Convergent evolution of complex brains, cognition and sleep functions in mammals and birds. *Neuroscience and Biobehavioral Reviews*, 2009, 33: 253-70.

Rechtschaffen, A. Current perspectives on the function of sleep. *Perspectives in Biology and Medicine*, 1998, 41: 359-90.

Rechtschaffen, A. and Kales, A. A manual of standardized terminology, techniques and scoring system of sleep stages in human subjects. In, Los Angeles: Brain Information Service/Brain Research Institute, University of California, 1968.

Reimann, M. W., Nolte, M., Scolamiero, M. *et al.* Cliques of Neurons Bound into Cavities Provide a Missing Link between Structure and Function. *Frontiers in Computational Neuroscience*, 2017, 11

Rey, M., Bastuji, H., Garcia-Larrea, L., Guillemant, P., Mauguiere, F. and Magnin, M. Human thalamic and cortical activities assessed by dimension of activation and spectral edge frequency during sleep wake cycles. *Sleep*, 2007, 30: 907-12.

Reynolds, C. F., Buysse, D. J. and Horne, J. Why we sleep: The functions of sleep in humans and other mammals. *American Journal of Psychiatry*, 1991, 148: 944-45.

Roach, B. and Mathalon, D. Event-related EEG time-frequency analysis: An overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia Bulletin*, 2008, 34: 907-26.

Rodriguez, A. V., Funk, C. M., Vyazovskiy, V. V., Nir, Y., Tononi, G. and Cirelli, C. Why Does Sleep Slow-Wave Activity Increase After Extended Wake? Assessing the Effects of Increased Cortical Firing During Wake and Sleep. *Journal of Neuroscience*, 2016, 36: 12436-47.

Rogers, R. D., Owen, A. M., Middleton, H. C. *et al.* Choosing between small, likely rewards and large, unlikely rewards activates inferior and orbital prefrontal cortex. *Journal of Neuroscience*, 1999, 19: 9029-38.

Roy, A. A theory of the brain: localist representation is used widely in the brain. *Front Psychol*, 2012, 3: 551.

Rusterholz, T., Durr, R. and Achermann, P. Inter-individual Differences in the Dynamics of Sleep Homeostasis. *Sleep*, 2010, 33: 491-98.

Sakai, K., Petitjean, F. and Jouvet, M. Effects of ponto-mesencephalic lesions and electrical stimulation upon PGO waves and EMPs in unanesthetized cats. *Electroencephalography and Clinical Neurophysiology*, 1976, 41: 49-63.

Saletin, J. M., Van Der Helm, E. and Walker, M. P. Structural brain correlates of human sleep oscillations. *Neuroimage*, 2013, 83: 658-68.

Salone, A., Di Giacinto, A., Lai, C. *et al.* The Interface between Neuroscience and Neuro-Psychoanalysis: Focus on Brain Connectivity. *Frontiers in Human Neuroscience*, 2016, 10

Sameshima, K., Baccala, L. A., Alvarenga, M. Y. and Yang, H. M. Nonlinear analysis of theta rhythm recorded during desynchronized sleep in rats. *Society for Neuroscience Abstracts*, 1997, 23: 1847-47.

Sameshima, K., Baccala, L. A., Ballester, G., Valle, A. C. and Timo-Iaria, C. Causality analysis of rhythmic activities of desynchronized sleep in the rat. *Society for Neuroscience Abstracts*, 1996, 22: 27-27.

Sameshima, K., Baccala, L. A., Ballester, G., Valle, A. C. and Timo-Iaria, C. Statistical testing in probing causality in the interaction of brain structures. *Society for Neuroscience Abstracts*, 1998, 24: 133-33.

Sarasso, S., Proserpio, P., Moroni, F. *et al.* Hippocampal sleep spindles preceding neocortical sleep onset in humans: an intracerebral EEG study. *Journal of Sleep Research*, 2012, 21: 130-30.

Sarasso, S., Proserpio, P., Pigorini, A. *et al.* Hippocampal sleep spindles preceding neocortical sleep onset in humans. *Neuroimage*, 2014, 86: 425-32.

Sarter, M., Bruno, J. P. and Givens, B. Attentional functions of cortical cholinergic inputs: What does it mean for learning and memory? *Neurobiology of Learning and Memory*, 2003, 80: 245-56.

Savage, V. M. and West, G. B. A quantitative, theoretical framework for understanding mammalian sleep. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104: 1051-56.

Schelter, B., Timmer, J. and Eichler, M. Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. *Journal of Neuroscience Methods*, 2009, 179: 121-30.

Schelter, B., Winterhalder, M., Eichler, M. *et al.* Testing for directed influences among neural signals using partial directed coherence. *Journal of Neuroscience Methods*, 2006, 152: 210-19.

Scheuler, W., Rappelsberger, P., Pastelakprice, C., Kubicki, S. and Petsche, H. Alpha-activity of NREM sleep. *Basic Mechanisms of the Eeg*, 1993: 183-214.

Scheuler, W., Stkubicki, Marquardt, J., Scholz, G., Weiss, K. H. and Gaeth, L. Spectral and coherence analysis of pharmacologically induced changes of the alpha sleep pattern. *Electroencephalography and Clinical Neurophysiology*, 1988, 69: P23-P23.

Schmidt, M., Swang, T., Hamilton, I. and Best, J. Sleep and energy conservation: a new paradigm based on state-dependent coupling of biological operations. *Journal of Sleep Research*, 2016, 25: 240-41.

Schmidt, M. H. The energy allocation function of sleep: A unifying theory of sleep, torpor, and continuous wakefulness. *Neuroscience and Biobehavioral Reviews*, 2014, 47: 122-53.

Schneider, B. and Koenigs, M. Human lesion studies of ventromedial prefrontal cortex. *Neuropsychologia*, 2017, 107: 84-93.

Seeck, M., Michel, C. M., Pascual-Marqui, R. D. *et al.* Localization and propagation of epileptic seizures using low resolution electromagnetic tomography (LORETA). *Brain Topography*, 1996, 8: 414-15.

Seeck-Hirschner, M., Baier, P. C., Weinhold, S. L. *et al.* Declarative Memory Performance Is Associated With the Number of Sleep Spindles in Elderly Women. *American Journal of Geriatric Psychiatry*, 2012, 20: 782-88.

Sena, P., D'amore, M., Brandimonte, M. A., Squitieri, R. and Fiorentino, A. Experimental framework for simulators to study driver cognitive distraction: brake reaction time in different levels of arousal. *Transport Research Arena Tra2016*, 2016, 14: 4410-19.

Seneviratne, U. Rational Manipulation of Digital EEG: Pearls and Pitfalls. *Journal of Clinical Neurophysiology*, 2014, 31: 507-16.

Seth, A., Barrett, A. and Barnett, L. Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, 2015, 35: 3293-97.

Seth, A. K., Barrett, A. B. and Barnett, L. Causal density and integrated information as measures of conscious level. *Philos Trans A Math Phys Eng Sci*, 2011, 369: 3748-67.

Shannon, C. A mathematical theory of communication. *Bell System Technical Journal*, 1948, 27: 623-56.

Shine, J. M., Koyejo, O., Bell, P. T., Gorgolewski, K. J., Gilat, M. and Poldrack, R. A. Estimation of dynamic functional connectivity using Multiplication of Temporal Derivatives. *Neuroimage*, 2015, 122: 399-407.

Siclari, F., Baird, B., Perogamvros, L. *et al.* The neural correlates of dreaming. *Nature Neuroscience*, 2017, 20: 872-+.

Siclari, F., Bernardi, G., Riedner, B. A., Larocque, J. J., Benca, R. M. and Tononi, G. Two Distinct Synchronization Processes in the Transition to Sleep: A High-Density Electroencephalographic Study. *Sleep*, 2014, 37: 1621-U214.

Siegel, J. M. Clues to the functions of mammalian sleep. *Nature*, 2005, 437: 1264-71.

Siegel, J. M. SLEEP - OPINION Sleep viewed as a state of adaptive inactivity. *Nature Reviews Neuroscience*, 2009, 10: 747-53.

Siegel, J. M. REM sleep: A biological and psychological paradox. *Sleep Medicine Reviews*, 2011, 15: 139-42.

Silfverhuth, M., Hintsala, H., Kortelainen, J. and Seppanen, T. Experimental comparison of connectivity measures with simulated EEG signals. *Medical & Biological Engineering & Computing*, 2012, 50: 683-88.

Smith, C. Sleep states, memory processes and synaptic plasticity. *Behavioural Brain Research*, 1996, 78: 49-56.

Smith, C. Sleep states and memory processes in humans: procedural versus declarative memory systems. *Sleep Medicine Reviews*, 2001, 5: 491-506.

Sommerlade, L., Amtage, F., Lapp, O. *et al.* On the estimation of the direction of information flow in networks of dynamical systems. *Journal of Neuroscience Methods*, 2011, 196: 182-89.

Sommerlade, L., Thiel, M., Platt, B. *et al.* Inference of Granger causal time-dependent influences in noisy multivariate time series. *Journal of Neuroscience Methods*, 2012, 203: 173-85.

Speth, C. and Speth, J. The borderlands of waking: Quantifying the transition from reflective thought to hallucination in sleep onset. *Consciousness and Cognition*, 2016, 41: 57-63.

Spoormaker, V. I., Schroter, M. S., Gleiser, P. M. *et al.* Development of a Large-Scale Functional Brain Network during Human Non-Rapid Eye Movement Sleep. *Journal of Neuroscience*, 2010, 30: 11379-87.

Sporns, O. and Betzel, R. F. Modular Brain Networks. *Annual Review of Psychology, Vol 67*, 2016, 67: 613-40.

Sporns, O., Honey, C. J. and Kotter, R. Identification and Classification of Hubs in Brain Networks. *Plos One*, 2007, 2

Stam, C. J., Van Straaten, E. C. W., Van Dellen, E. *et al.* The relation between structural and functional connectivity patterns in complex brain networks. *International Journal of Psychophysiology*, 2016, 103: 149-60.

Stickgold, R. Sleep-dependent memory consolidation. *Nature*, 2005, 437: 1272-78.

Stickgold, R. Parsing the role of sleep in memory processing. *Current Opinion in Neurobiology*, 2013, 23: 847-53.

Stickgold, R. and Walker, M. P. Memory consolidation and reconsolidation: what is the role of sleep? *Trends in Neurosciences*, 2005a, 28: 408-15.

Stickgold, R. and Walker, M. P. Sleep and memory: The ongoing debate. *Sleep*, 2005b, 28: 1225-27.

Sämann, P. G., Tully, C., Spoormaker, V. I. *et al.* Increased sleep pressure reduces resting state functional connectivity. *Magnetic Resonance Materials in Physics Biology and Medicine*, 2010, 23: 375-89.

Sämann, P. G., Wehrle, R., Hoehn, D. *et al.* Development of the Brain's Default Mode Network from Wakefulness to Slow Wave Sleep. *Cerebral Cortex*, 2011, 21: 2082-93.

Tass, P., Rosenblum, M. G., Weule, J. *et al.* Detection of n : m phase locking from noisy data: Application to magnetoencephalography. *Physical Review Letters*, 1998, 81: 3291-94.

Tobler, I. Is sleep fundamentally different between mammalian species. *Behavioural Brain Research*, 1995, 69: 35-41.

Toien, O., Blake, J. and Barnes, B. M. Thermoregulation and energetics in hibernating black bears: metabolic rate and the mystery of multi-day body temperature cycles. *Journal of Comparative Physiology B-Biochemical Systemic and Environmental Physiology*, 2015, 185: 447-61.

Tononi, G., Boly, M., Massimini, M. and Koch, C. Integrated information theory: from consciousness to its physical substrate. *Nature Reviews Neuroscience*, 2016, 17: 450-61.

Tononi, G. and Cirelli, C. Steep function and synaptic homeostasis. *Sleep Medicine Reviews*, 2006, 10: 49-62.

Tononi, G. and Cirelli, C. Sleep and the Price of Plasticity: From Synaptic and Cellular Homeostasis to Memory Consolidation and Integration. *Neuron*, 2014, 81: 12-34.

Trinder, M., Bisanz, J. E., Burton, J. P. and Reid, G. Bacteria Need "Sleep" Too?: Microbiome Circadian Rhythmicity, Metabolic Disease, and Beyond. *University of Toronto Medical Journal*, 2015, 92: 52-55.

Tsoukalas, I. The Origin of REM Sleep: A Hypothesis. *Dreaming*, 2012, 22: 253-83.

Tung, A., Takase, L., Fornal, C. and Jacobs, B. Effects of sleep deprivation and recovery sleep upon cell proliferation in adult rat dentate gyrus. *Neuroscience*, 2005, 134: 721-23.

Uehara, T., Yamasaki, T., Okamoto, T. *et al.* Efficiency of a "Small-World" Brain Network Depends on Consciousness Level: A Resting-State fMRI Study. *Cerebral Cortex*, 2014, 24: 1529-39.

Ulrich, M., Keller, J., Hoenig, K., Waller, C. and Gron, G. Neural correlates of experimentally induced flow experiences. *Neuroimage*, 2014, 86: 194-202.

Valderrama, M., Crepon, B., Botella-Soler, V. *et al.* Human Gamma Oscillations during Slow Wave Sleep. *Plos One*, 2012, 7

Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J. and Friston, K. J. Effective connectivity: Influence, causality and biophysical modeling. *Neuroimage*, 2011, 58: 339-61.

Valli, K., Revonsuo, A., Palkas, O., Ismail, K. H., Ali, K. J. and Punamaki, R. L. The threat simulation theory of the evolutionary function of dreaming: Evidence from dreams of traumatized children. *Consciousness and Cognition*, 2005, 14: 188-218.

Van Der Helm, E., Yao, J., Dutt, S., Rao, V., Saletin, J. M. and Walker, M. P. REM Sleep Depotentiates Amygdala Activity to Previous Emotional Experiences. *Current Biology*, 2011, 21: 2029-32.

Van Vugt, M. K., Sederberg, P. B. and Kahana, M. J. Comparison of spectral analysis methods for characterizing brain oscillations. *Journal of Neuroscience Methods*, 2007, 162: 49-63.

Vanoosterom, A. History and evolution of methods for solving the inverse problem. *Journal of Clinical Neurophysiology*, 1991, 8: 371-80.

Vassalli, A. and Dijk, D. J. Sleep function: current questions and new approaches. *European Journal of Neuroscience*, 2009, 29: 1830-41.

Vecchio, F., Miraglia, F., Gorgoni, M. *et al.* Cortical Connectivity Modulation During Sleep Onset: A Study via Graph Theory on EEG Data. *Human Brain Mapping*, 2017, 38: 5456-64.

Vicente, R., Wibral, M., Lindner, M. and Pipa, G. Transfer entropy-a model-free measure of effective connectivity for the neurosciences. *Journal of Computational Neuroscience*, 2011, 30: 45-67.

Vogt, B. A. and Laureys, S. Posterior cingulate, precuneal and retrosplenial cortices: cytology and components of the neural network correlates of consciousness. *Boundaries of Consciousness: Neurobiology and Neuropathology*, 2005, 150: 205-17.

Waberski, T. D., Kreitschmann-Andermahr, I., Kawohl, W. *et al.* Spatio-temporal source imaging reveals subcomponents of the human auditory mismatch negativity in the cingulum and right inferior temporal gyrus. *Neuroscience Letters*, 2001, 308: 107-10.

Wan, X. H., Sekiguchi, A., Yokoyama, S., Riera, J. and Kawashima, R. Electromagnetic source imaging: Backus-Gilbert resolution spread function-constrained and functional MRI-guided spatial filtering. *Human Brain Mapping*, 2008, 29: 627-43.

Wanguemert-Perez, J. G., Godoy-Rubio, R., Ortega-Monux, A. and Molina-Fernandez, I. Removal of the Gibbs phenomenon and its application to fast-Fourier-transform-based mode solvers. *Journal of the Optical Society of America a-Optics Image Science and Vision*, 2007, 24: 3772-80.

Watanabe, T. Hypnagogic imagery: Relationship to mental imagery, personality, and creativity. *International Journal of Psychology*, 1996, 31: 38454-54.

Weiss, S. A., Mckhann, G., Goodman, R. *et al.* Field effects and ictal synchronization: insights from in homine observations. *Frontiers in Human Neuroscience*, 2013, 7

Werth, E., Achermann, P. and Borbely, A. A. Fronto-occipital EEG power gradients in human sleep. *Journal of Sleep Research*, 1997, 6: 102-12.

Wibral, M., Vicente, R. and Lindner, M. Transfer Entropy in Neuroscience. In: M. WIBRAL, R. VICENTE and J. T. LIZIER (Eds), *Directed Information Measures in Neuroscience*, 2014: 3-36.

Wollstadt, P., Martinez-Zarzuela, M., Vicente, R., Diaz-Pernas, F. J. and Wibral, M. Efficient Transfer Entropy Analysis of Non-Stationary Neural Time Series. *Plos One*, 2014, 9

Wright, K. P., Badia, P. and Wauquier, A. Topographical and temporal patterns of brain activity during the transition from wakefulness to sleep. *Sleep*, 1995, 18: 880-89.

Wu, W., Jia, W. Y., Liu, H. S., Gao, X. R., Zhang, G. J. and Wang, Y. P. Localization of epileptic foci from preictal EEG data using standardized shrinking LORETA-FOCUSS algorithm. *2005 First International Conference on Neural Interface and Control Proceedings*, 2005: 209-12.

Xie, L. L., Kang, H. Y., Xu, Q. W. *et al.* Sleep Drives Metabolite Clearance from the Adult Brain. *Science*, 2013, 342: 373-77.

Zeitlhofer, J., Gruber, G., Anderer, P., Asenbaum, S., Schimicek, P. and Saletu, B. Topographic distribution of sleep spindles in young healthy subjects. *Journal of Sleep Research*, 1997, 6: 149-55.

Zoetmulder, M., Nikolic, M., Biernat, H., Korbo, L., Friberg, L. and Jennum, P. Increased Motor Activity During REM Sleep Is Linked with Dopamine Function in Idiopathic REM Sleep Behavior Disorder and Parkinson Disease. *Journal of Clinical Sleep Medicine*, 2016, 12: 895-903.

## Quotations

0. https://simple.wikiquote.org/wiki/William_Shakespeare

1.1. https://www.brainyquote.com/quotes/arthur_schopenhauer_118678

1.2. https://www.goodreads.com/quotes/26601-each-night-when-i-go-to-sleep-i-die-and

1.3. https://simple.wikiquote.org/wiki/William_Shakespeare

1.4. http://ngm.nationalgeographic.com/2010/05/sleep/max-text

1.5. https://www.brainyquote.com/quotes/william_blake_150142

1.6. https://www.brainyquote.com/quotes/thomas_dekker_204715

1.7. https://www.nature.com/articles/nature04286

1.8. https://www.shmoop.com/odyssey/book-xi-quotes-3.html

1.9. http://www.azquotes.com/quote/992907

2.1. https://www.goodreads.com/quotes/tag/connection

2.2. https://arxiv.org/ftp/arxiv/papers/0710/0710.3341.pdf (page 20)

2.3. https://www.livescience.com/46511-understanding-sleep-with-mathematics-nsf-bts.html

3.1. https://www.goodreads.com/quotes/8727562-on-the-walls-of-the-cave-only-the-shadows-are

4.1. https://www.goodreads.com/quotes/361785-if-you-want-to-find-the-secrets-of-the-universe

5.1. https://www.brainyquote.com/quotes/shakuntala_devi_598202?src=t_mathematics

6.1. https://www.goodreads.com/quotes/962093-science-is-more-than-a-body-of-knowledge-it-is

6.2. https://www.idrlabs.com/intp.php

6.3. https://www.goodreads.com/quotes/1195886-if-a-solution-fails-to-appear-and-yet-we

6.4. http://www.lexiyoga.com/sleep-quotes

6.5. http://www.lexiyoga.com/sleep-quotes

6.6. https://www.brainyquote.com/quotes/friedrich_nietzsche_13460

# ANTONIO FERNANDEZ GUERRERO

## Summary

**• Theoretical modelling with advanced mathematics. Bachelor in theoretical physics, master in biophysics, doctoral studies in neuroscience.**
**• Extraction of hidden patterns in neurobiological signal processing, including causal and directional components (besides simple correlations).**
**• Broad background in physics and neuroscience, focused on how best interpreting complex quantitative data.**

*Name*: ANTONIO FERNANDEZ GUERRERO.
*Birth*: 5th of October, 1987 (Málaga, Spain).
*Nationality*: Spanish
*Telef*: 0786726499.
*E-mail*: antoniofernandezguerrero@hotmail.es or a.fernandez@pharma.uzh.ch.

## Core Competencies

- Advanced and strong background in mathematics, physics, statistics and data mining.
- Application of programs (LORETA and Matlab) for studies of signal processing analysis in the brain.
- Advanced Matlab programmer and LORETA user.
- Constantly expanding into new scientific and intellectual horizons.
- Analytical thinker and problem-solver.

## Education and Professional Experience

**2003-2005**: Bachillerato (gymnasium) studies specialized in sciences in private center “Sagrado Corazón” (Málaga).

**2005-2011**: Bachelor studies. Complutense University of Madrid.

- Physics degree was structured to offer several paths (e.g., astrophysics, geophysics, etc.). I chose theoretical physics, arguably the one with highest level of abstraction and advanced mathematics.
- Honors in 21 subjects. Final bachelor's average: 8.70/10.

**2011-2012**: Master studies. Complutense University/Cajal Blue Brain Project

- My master was conducted in the field of biomedical physics. I pursued the most theoretical and mathematical branch inside the master: biophysics and relationships to brain signal processing.
- Title of master thesis (translated to English): “Analysis of magnetoencephalographic signals applied to the study of mild cognitive impairment. Synchronization and directionality”.
- Master thesis was conducted in collaboration with Cajal Blue Brain Project (spanish branch of Human Brain Project, which provided the

data and supervision). The topic was statistical analyses of disruption in brain signals describing cognitive deficits appearing in elderly population affected by "mild cognitive impairment". These analyses combined both correlative and directionality aspects using modern mathematical tools (Rosenblum and Pikovsky's method and phase-locking value). All necessary programming was done entirely by myself (Matlab).

- Master thesis research was assessed at the end of the master by a scientific committee, granting a grade of 9 (with 10 being the maximum grade).
- Honors in 3 subjects. Final master's average: 9.09/10.

**2012-2013**: Assistant in office of economics.

Before starting the PhD in Zürich, I worked in Málaga in the office of Antonio Fernández Benitez (economist, my father) as assistant for a period between half a year and a whole year, aiding with graph analyses and other data processing (using Excel and other programs).

**2013 (starting 1st of November)-2018**: Doctoral studies. University of Zürich (Institute of Pharmacology and Toxicology)/ZNZ

- My doctoral studies were carried on the field of computational neuroscience, applied to the problem of human sleep and its transitions, under the supervision of prof. Peter Achermann. This research project applied isolated effective coherence and power spectral analyses to track the evolution of sleep onset, from vigilance to deep sleep, in conjunction to a well-validated source localization method (eLORETA).
- In parallel to my PhD studies in Neuroscience, I took time to study many courses of my interest on the online platform "Coursera" (with special emphasis on neuroscience and psychology) and in the University of Zürich (e.g., ethics and german).

## Congresses, seminars and workshops

Neurex symposium in Basel (11/8/2013).
Swiss society for sleep research, sleep medicine and chronobiology (SSSSC) meeting in Luzern (5/15-16/2014).
Zürich center for integrative human physiology (ZIHP) seminar (10/7/2014).
Clinical research priority program (CRPP) sleep and health workshops (1/15-17/2015).
ZIHP seminar (4/14/2015).
PhD retreat in Valens (5/7-9/2015).
SSSSC congress in Interlaken (6/11-12/2015).
CRPP meeting (7/3/2015).
ZIHP seminar (8/21/2015),
LORETA workshop (9/4/2015).
Neuroscience Center Zürich (ZNZ) symposium (9/11/2015).
CRPP meeting (1/8/2016).
CRPP retreat at Epi Clinic (2/5/2016).
ZIHP seminar (3/18/2016).
CRPP meeting (3/31 and 4/1/2016).
LORETA course (7/4-5/2016).
CRPP meeting (7/8/2016).
ZIHP seminar (8/26/2016).
ESRS congress in Bologna (9/20-23/2016).
CRPP meeting (10/7/2016).
ZIHP seminar (4/20/2017).
CRPP meeting (5/19/2017).
Connectivity workshop in Bern (8/30/2017).
ZIHP meeting (9/1/2017).
CRPP retreat at Epi Clinic (9/15/2017).

## Skills and hobbies

- Languages: spanish (native), english (C1), german (A2.2).
- Studies of piano for 7 years. Love for classical music (especially Mozart and Chopin), jazz and rock.
- Reading books and watching documentaries on psychology & neurosciences, philosophy and history.
- Autodidact. Lifelong adherence to the doctrine: "Self-education is, I firmly believe, the only kind of education there is" (Isaac Asimov).
- Antonio is member of Mensa (international association for intellectually gifted adults; membership number: E-02847).
- Antonio enjoys hiking and being in contact with nature, both in Spain and Switzerland.
- Antonio enjoys watching complex TV series (e.g., "House", “Genius”, and "Game of Thrones") or movies (preferred genres are science fiction and all sorts of thrillers).

## Publications

Omlin, X., Crivelli, M., Fernandez Guerrero, A. et al. The Effect of a Slowly Rocking Bed on Sleep. Scientific Reports, 8 (2018): 2156.

Blanco Perez, C. and Fernandez Guerrero, A. Expanding Universe with a Variable Cosmological Term. Zeitschrift Für Naturforschung Section a-a Journal of Physical Sciences, 2015, 70: 905-11.